



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|-----------|---|
| (51) International Patent Classification ⁶ : A61K 38/00, 39/00, 39/395, 39/40, C07H 21/04, C07K 1/00, C12N 15/00, C12P 21/06, G01N 33/53 | A1 | (11) International Publication Number: WO 98/19689 (43) International Publication Date: 14 May 1998 (14.05.98) |
| (21) International Application Number: PCT/US97/19226 (22) International Filing Date: 27 October 1997 (27.10.97) (30) Priority Data: 60/029,930 1 November 1996 (01.11.96) US (71) Applicants (for all designated States except US): SMITHK-LINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHK-LINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): BLACK, Michael, Terance [GB/US]; 502 Milhouse Way, Chester Springs, PA 19425 (US). HODGSON, John, Edward [GB/US]; 260 Lapp Road, Malvern, PA 19355 (US). KNOWLES, David, Justin, Charles [GB/GB]; Downsview House, 45 Cronks Hill Road, Redhill, Surrey RH1 6LY (GB). LONETTO, Michael, Arthur [US/US]; 18 Victoria Circle, Collegeville, PA 19426 (US). NICHOLAS, Richard, Oakley [GB/US]; 355 Carmen Drive, Collegeville, PA 19426 (US). REID, Robert, H., Jr. [US/US]; 8 Pacer Lane, East Norriton, PA 19401 (US). | | ZARFOS, Phillip, N. [US/US]; 1907 Yorktown North, Norristown, PA 19403 (US). (74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US). (81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> |
| (54) Title: NOVEL CODING SEQUENCES (57) Abstract <p>This invention relates to newly identified Streptococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.</p> | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakhstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

NOVEL CODING SEQUENCES

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides set forth in Table 1.

BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including otitis media, pneumonia and meningitis. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* (herein *S. pneumoniae*) has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *S. pneumoniae*, many questions concerning the virulence of this microbe remain.

While certain Streptococcal factors associated with pathogenicity have been identified, e.g., capsule polysaccharides, peptidoglycans, pneumolysins, PspA Complement factor H binding component, autolysin, neuraminidase, peptide permeases, hydrogen peroxide, IgA1 protease, the list is certainly not complete. Further very little is known concerning the temporal expression of such genes during infection and disease progression in a mammalian host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis. In addition to providing a fuller understanding of known proteins, such an approach will identify previously unrecognised targets.

GUG is used as an initiating nucleotide, rather than ATG, for a significant number of mRNA's in both Gram positive and Gram negative bacteria. Statistics on the frequency of NTG codons in the start codon for several bacterial species are available on line via computer at http://biochem.otago.ac.nz:800/Transterm/home_page.html).

A discussion of initiation codons in *B. subtilis* is set forth in Vellanoweth, RL.1993 in Bacillus subtilis and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711. Vellenworth indicates a major difference between *B. subtilis* and the gram-negative organisms is in the choice of initiation codon. 91% of the sequenced *E. coli*

genes start with AUG. By contrast, about 30% of *B. subtilis* and other clostridial branch — genes start with UUG or GUG. Moreover, CUG functions as a start codon in *B. subtilis*. Mutations of an AUG initiation codon to GUG or UUG often cause decreased expression in *B. subtilis* and *E. coli*. Generally, translation efficiency is higher with AUG initiation codons. A strong Shine-Delgarno ribosome binding site, however, can compensate almost fully for a weak initiation codon. It has been reported that genes with a range of expression levels have initiation codons other than ATG in gram positives (Vellanoweth, RL.1993 in Bacillus subtilis and other Gram Positive Bacteria. Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711).

Provided herein are ORF sequences from genes possessing GUG initiation codons and proteins expressed therefrom and homologues thereto to be used for screening for antimicrobial compounds. Clearly, there is a need for polypeptide and polynucleotide sequences that may be used to screen for antimicrobial compound and which may also be used to determine the roles of such sequences in pathogenesis of infection, dysfunction and disease. There is also need, therefore, for identification and characterization of such sequences which may play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known protein(s) as set forth in Table 1.

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel polypeptides by homology between an amino acid sequence selected from the group consisting of the sequences set out in Table 1 and a known amino acid sequence or sequences of other proteins such as the protein identities listed in Table 1.

It is a further object of the invention to provide polynucleotides that encode novel polypeptides, particularly polynucleotides that encode polypeptides of *Streptococcus pneumoniae*.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding a polypeptide comprising a sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In another particularly preferred embodiment of the invention there is a novel protein from *Streptococcus pneumoniae* comprising an amino acid sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding a polypeptide of the invention, particularly *Streptococcus pneumoniae* polypeptide, and including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of a polypeptide of the invention and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of the polypeptides of the invention encoded by naturally occurring alleles of their genes.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing expression of the polypeptides and polynucleotides of the invention, treating disease, for example, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, assaying genetic variation, and administering a polypeptide or polynucleotide of the invention to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to a polynucleotide sequence of the invention, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against polypeptides of the invention.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided agonists and antagonists of the polypeptides and polynucleotides of the invention, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a polynucleotide or a polypeptide of the invention for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means any bacterial infection, but preferably a streptococcal infection, such as, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema, endocarditis, meningitis, and infection of cerebrospinal fluid.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings

of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the tested polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the test amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino

acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for

example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides — often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

Each of polynucleotide and polypeptide sequences provided herein may be used in the discovery and development of antibacterial compounds. Upon expression of the sequences with the appropriate initiation and termination codons the encoded polypeptide can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding preferably the amino terminal regions of the encoded protein or the Shine-Delgarno region can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The

first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. Because each of the sequences contains an open reading frame (ORF) with an appropriate initiation and termination codons, the encoded protein upon expression can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

It is believed that bacteria possess a number of ways of regulating gene expression levels, especially in subtle degrees, and the interplay between ribosome binding site and initiation codon is utilized for this purpose for these genes. It is also believed that such genes will be important targets for antimicrobial drug discovery, particularly since pathogenesis genes are believed undergo gene expression regulation during in the pathogenesis process. Therefore, the invention provides ORF sequences possessing a GTG (GUG) initiation codon and protein targets expressed therefrom.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

ORF Gene Expression

Recently techniques have become available to evaluate temporal gene expression in bacteria, particularly as it applies to viability under laboratory and infection conditions. A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of an ORF unknown by one of these methods yields additional information about its function and permits the selection of such an ORF for further development as a screening target. Briefly, these approaches include:

1) **Signature Tagged Mutagenesis (STM):** This technique is described by Hensel *et al.*, Science 269: 400-403(1995), the contents of which is incorporated by reference for background purposes. Signature tagged mutagenesis identifies genes necessary for the establishment/maintenance of infection in a given infection model.

The basis of the technique is the random mutagenesis of target organism by various means (e.g., transposons) such that unique DNA sequence tags are inserted in close proximity to the site of mutation. The tags from a mixed population of bacterial mutants and bacteria recovered from an infected hosts are detected by amplification, radiolabeling and hybridisation analysis. Mutants attenuated in virulence are revealed by absence of the tag from the pool of bacteria recovered from infected hosts.

In *Streptococcus pneumoniae*, because the transposon system is less well developed, a more efficient way of creating the tagged mutants is to use the insertion-duplication mutagenesis technique as described by Morrison *et al.*, J. Bacteriol. 159:870 (1984) the contents of which is incorporated by reference for background purposes.

2) **In Vivo Expression Technology (IVET):** This technique is described by Camilli *et al.*, Proc. Nat'l. Acad. Sci. USA. 91:2634-2638 (1994), the contents of which is incorporated by reference for background purposes. IVET identifies genes up-regulated during infection when compared to laboratory cultivation, implying an important role in infection. ORF identified by this technique are implied to have a significant role in infection establishment/maintenance.

In this technique random chromosomal fragments of target organism are cloned upstream of a promoter-less recombinase gene in a plasmid vector. This construct is introduced into the target organism which carries an antibiotic resistance gene flanked by resolvase sites. Growth in the presence of the antibiotic removes from the population those fragments cloned into the plasmid vector capable of supporting transcription of the recombinase gene and therefore have caused loss of antibiotic resistance. The resistant pool

is introduced into a host and at various times after infection bacteria may be recovered and assessed for the presence of antibiotic resistance. The chromosomal fragment carried by each antibiotic sensitive bacterium should carry a promoter or portion of a gene normally upregulated during infection. Sequencing upstream of the recombinase gene allows identification of the up regulated gene.

3) Differential display: This technique is described by Chuang *et al.*, *J. Bacteriol.* 175:2026-2036 (1993), the contents of which is incorporated by reference for background purposes. This method identifies those genes which are expressed in an organism by identifying mRNA present using randomly-primed RT-PCR. By comparing pre-infection and post infection profiles, genes up and down regulated during infection can be identified and the RT-PCR product sequenced and matched to ORF 'unknowns'.

4) Generation of conditional lethal mutants by transposon mutagenesis: This technique, described by de Lorenzo, V. *et al.*, *Gene* 123:17-24 (1993); Neuwald, A. F. *et al.*, *Gene* 125: 69-73(1993); and Takiff, H. E. *et al.*, *J. Bacteriol.* 174:1544-1553(1992), the contents of which is incorporated by reference for background purposes, identifies genes whose expression are essential for cell viability.

In this technique transposons carrying controllable promoters, which provide transcription outward from the transposon in one or both directions, are generated. Random insertion of these transposons into target organisms and subsequent isolation of insertion mutants in the presence of inducer of promoter activity ensures that insertions which separate promoter from coding region of a gene whose expression is essential for cell viability will be recovered. Subsequent replica plating in the absence of inducer identifies such insertions, since they fail to survive. Sequencing of the flanking regions of the transposon allows identification of site of insertion and identification of the gene disrupted. Close monitoring of the changes in cellular processes/morphology during growth in the absence of inducer yields information on likely function of the gene. Such monitoring could include flow cytometry (cell division, lysis, redox potential, DNA replication), incorporation of radiochemically labeled precursors into DNA, RNA, protein, lipid, peptidoglycan, monitoring reporter enzyme gene fusions which respond to known cellular stresses.

5) Generation of conditional lethal mutants by chemical mutagenesis: This technique is described by Beckwith, J., *Methods in Enzymology* 204: 3-18(1991), the contents of which are incorporated herein by reference for background purposes. In this technique random chemical mutagenesis of target organism, growth at

temperature other than physiological temperature (permissive temperature) and subsequent replica plating and growth at different temperature (e.g. 42°C to identify ts, 25°C to identify cs) are used to identify those isolates which now fail to grow (conditional mutants). As above close monitoring of the changes upon growth at the non-permissive temperature yields information on the function of the mutated gene. Complementation of conditional lethal mutation by library from target organism and sequencing of complementing gene allows matching with unknown ORF.

6) RT-PCR: *Streptococcus pneumoniae* messenger RNA is isolated from bacterial infected tissue e.g. 48 hour murine lung infections, and the amount of each mRNA species assessed by reverse transcription of the RNA sample primed with random hexanucleotides followed by PCR with gene specific primer pairs. The determination of the presence and amount of a particular mRNA species by quantification of the resultant PCR product provides information on the bacterial genes which are transcribed in the infected tissue. Analysis of gene transcription can be carried out at different times of infection to gain a detailed knowledge of gene regulation in bacterial pathogenesis allowing for a clearer understanding of which gene products represent targets for screens for novel antibacterials. Because of the gene specific nature of the PCR primers employed it should be understood that the bacterial mRNA preparation need not be free of mammalian RNA. This allows the investigator to carry out a simple and quick RNA preparation from infected tissue to obtain bacterial mRNA species which are very short lived in the bacterium (in the order of 2 minute halflives). Optimally the bacterial mRNA is prepared from infected murine lung tissue by mechanical disruption in the presence of TRIzol (GIBCO-BRL) for very short periods of time, subsequent processing according to the manufacturers of TRIzol reagent and DNAase treatment to remove contaminating DNA. Preferably the process is optimised by finding those conditions which give a maximum amount of *Streptococcus pneumoniae* 16S ribosomal RNA as detected by probing Northern blots with a suitably labelled sequence specific oligonucleotide probe. Typically a 5' dye labelled primer is used in each PCR primer pair in a PCR reaction which is terminated optimally between 8 and 25 cycles. The PCR products are separated on 6% polyacrylamide gels with detection and quantification using GeneScanner (manufactured by ABI).

Each of these techniques may have advantages or disadvantage depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind.

Use of the of these technologies when applied to the ORFs of the present invention enables identification of bacterial proteins expressed during infection, inhibitors of which would have utility in anti-bacterial therapy.

The invention relates to novel polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to known polypeptide as set forth in Table 1. The invention relates especially to compounds having the nucleotide and amino acid sequence selected from the group consisting of the sequences set out in Table 1, and to the nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

Deposited materials

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A deposit containing a *Streptococcus pneumoniae* bacterial strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned NCIMB Deposit No. 40794. The *Streptococcus pneumoniae* bacterial strain deposit is referred to herein as "the deposited bacterial strain" or as "the DNA of the deposited bacterial strain."

The deposited material is a bacterial strain that contains the full length FabH DNA, referred to as "NCIMB 40794" upon deposit.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

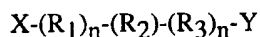
A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The deposited strain contains the full length genes comprising the polynucleotides set forth in Table 1. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

Polypeptides

The polypeptides of the invention include the polypeptides set forth in Table 1 (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of a polypeptide of the invention, and also those which have at least 50%, 60% or 70% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1 or the relevant portion, preferably at least 80% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 2000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in Table 1. In the formula above R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In preferred embodiments n is an integer between 1 and 1000 or 2000.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1, or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil

and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of polypeptides of the invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" is also used. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

Polynucleotides

The nucleotide sequences disclosed herein can be obtained by synthetic chemical techniques known in the art or can be obtained from *S. pneumoniae* 0100993 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers in a process of PCR-based cloning of the sequence from a bacterial genomic source. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

To obtain the polynucleotide encoding the protein using the DNA sequence given herein typically a library of clones of chromosomal DNA of *S.pneumoniae* 0100993 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by

Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory (see: Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70).

Moreover, another aspect of the invention relates to isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group consisting of the sequences in Table 1 and polynucleotides closely related thereto and variants thereof.

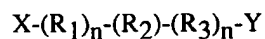
Using the information provided herein, such as the polynucleotide sequences set out in Table 1, a polynucleotide of the invention encoding polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence set forth in Table 1, typically a library of clones of chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotides set out in Table 1 were discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

The DNA sequences set out in Table 1 each contains at least one open reading frame encoding a protein having at least about the number of amino acid residues set forth in Table 1. The start and stop codons of each open reading frame (herein "ORF") DNA are the first three and the last three nucleotides of each polynucleotide set forth in Table 1.

Certain polynucleotides and polypeptides of the invention are structurally related to known proteins as set forth in Table 1. These proteins exhibit greatest homology to the homologue listed in Table 1 from among the known proteins.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence in Table 1. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R_1 and R_3 is any nucleic acid residue, n is an integer between 1 and 3000, and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in Table 1. In the polynucleotide formula above R_2 is oriented so that its 5' end residue is at the left, bound to R_1 , and its 3' end residue is at the right, bound to R_3 . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment n is an integer between 1 and 1000, or 2000 or 3000.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* having an amino acid sequence set out in Table 1. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1.

Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding polypeptide variants, that have the amino acid sequence of a polypeptide of Table 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of such polynucleotide.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding a polypeptide having the amino acid sequence set out in Table 1, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

A preferred embodiment is an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of: a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium

citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in Table 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding a polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in Table 1. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of each gene that comprises or is comprised by a polynucleotide set forth in Table 1 may be isolated by screening using a DNA sequence provided in Table 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the a polynucleotide or polypeptide sequence set forth in Table 1 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to

the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium

phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the polynucleotides of the invention for use as diagnostic reagents. Detection of such polynucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising a gene of the invention may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled polynucleotide sequences of the invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, *e.g.*, Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding a polypeptide of the invention can be used to identify and analyze mutations. These primers may be used for, among other things, amplifying a DNA of the invention isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably disease, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1. Increased or decreased expression of a polynucleotide of the invention can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of a polypeptide of the invention compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing recognition of a polypeptide of

the invention or from naive libraries (McCafferty, J. et al., (1990), *Nature* **348**, 552-554; Marks, J. et al., (1992) *Biotechnology* **10**, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) *Nature* **352**, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against a polypeptide of the invention may be employed to treat disease.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. et al. (1986), *Nature* **321**, 522-525 or Tempest et al., (1991) *Biotechnology* **9**, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into

muscles (Wolff et al., Hum Mol Genet 1992, 1:363, Manthorpe et al., Hum. Gene Ther. 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., J Biol Chem. 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, PNAS, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., Science 1989:243,375), particle bombardment (Tang et al., Nature 1992, 356:152, Eisenbraun et al., DNA Cell Biol 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., PNAS 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of a polypeptides or polynucleotides of the invention, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising a polypeptide of the invention and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be an agonist or antagonist of a polypeptide of the invention. The ability of the candidate molecule to agonize or antagonize a polypeptide of the invention is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of a polypeptide of the invention are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for antagonists of polypeptides of the invention is a competitive assay that combines any such polypeptide and a potential antagonist with a compound which binds such polypeptide, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. A polypeptide of the

invention can be labeled, such as by radioactivity or a colorimetric compound, such that the number of such polypeptide molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing activities induced by a polypeptide of the invention, thereby preventing the action of such polypeptide by excluding it from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of a polypeptide of the invention.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat disease.

Helicobacter pylori (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and Helicobacter Pylori (International Agency for Research on Cancer, Lyon, France; <http://www.uicc.ch/ecp/ecp2904.htm>). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with a polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of a polynucleotide or polypeptide of the invention, or a fragment or a variant thereof, for expressing such polynucleotide or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a polynucleotide of the

invention or protein coded therefrom, wherein the composition comprises a recombinant polynucleotide or protein coded therefrom comprising DNA which codes for and expresses an antigen of said polynucleotide or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A polypeptide of the invention or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain protein, such as, for example, those set forth in Table 1, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

Many orthopedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1 μ g/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

TABLES

Certain pertinent data for preferred polypeptide and polynucleotide embodiments of the invention are summarized in Tables 1 and 2.

Provided in Table 1 are sequence search results providing characterization information regarding certain preferred polynucleotides (denoted as "Assembly") and polypeptides of the invention encoded thereby. For each polynucleotide in Table 1, there is listed the closest homologue of each polypeptide encoded by each ORF in such polynucleotide. This determination of homology is based on a comparison of the sequences of in Table 1 with sequences available in the public domain (see heading entitled "Description" for the homologue name). Where no significant homologue was detected the term "unknown" appears after the heading "Description". Preferred polypeptides encoded by the ORFs of the invention, particularly full length proteins either obtained using such

ORFs or encoded entirely by such ORFs, are ones that have a biological function of the homologue listed, among other functions. The analysis used to determine each homologue listed in Table 1 was either BlastP and/or BlastX and/or MPSearch, each of which is well known. Also provided in Table 1 is the amino acid sequence encoded by each ORF. An "Assembly ID" number provides a convenient way to correlate the polynucleotide sequence with the ORF or ORFs it comprises and the polypeptides encoded by these ORFs, as well as to correlate such sequences with other pertinent information provided in Tables 1 and 2. Following the heading "ORF Predictions" the nucleotides at the beginning and end of the ORF sequence are set forth ("Start" and "End" respectively). The direction of translation on the polynucleotide depicted is denoted by an "F" for forward or an "R" for reverse (reverse being translated on the opposite strand from the one depicted). The length of each amino acid sequence is also indicated in a column entitled "Length." Below these data is shown the amino acid sequence encoded by the ORF. If a given polynucleotide comprises one ORF, then in the column entitled "ORF #" there is the numeral one. If it encodes two, there are the numerals one and two in the column, and so on.

TABLE 1

Assembly ID: 3049156

Assembly Length: 495bp

> 3049156 Strep Assembly -- Assembly id#3049156

```
CTCGGTGATAGAAATAGTGTAAATCATGCTTTTCTCTTCTTATCTATACTTTGCTACTTCT
ATTATACAAAAAATAAGCGCTTGACTAGGGATTTTTAGAAAAAAGCCTATTTTTTCA
AGAAAAATAGGCTTTTTCGAACGATTGACACAATTGGATTTGGTTAATTCACCTCTTAAC
GATGGTTTTTAAACGATATATATTTTTTATATATGTAAATTAAAACTTCTTTTCCTTTCACT
TCCTACGACTTTTCAGATACAGATAGCCAAAGAAGTTTTTCATAGAGGGCAAAAAAGAGGA
GGAAGGCATGAAGAAAGAAGGTCTCTGGCAAAATCATAATAACAGGATCCTTGGCTGGAT
CAAAAAGCCAGGTATCATCTCCCACAAAGAGAATTTGATGGAAAAGAGTAAAGAATTGGT
CAAAACCAATCAAACTCCCCCAAGTCCATCATCACAGGTAAGACTACTAGAGCCAGGAG
ACTTTTTTCGATAAAG
```

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-------|-----------|--------|
| ----- | ----- | ----- | ----- | ----- |
| 1 | 236 | 385 | R | 50 aa |

> 3049156-1 ORF translation from 236-385, direction R
VGDDTWLFDPAKDPVIMILPETFFLHAFLLFFALYENFFGYLYLKSRRK*

Description:
unknown

Assembly ID: 3049862
Assembly Length: 529bp

> 3049862 Strep Assembly -- Assembly id#3049862
CTAGAGCAAGTATTTTCAAACCTTTTCCGAATAAATAGATAGAGCCAGAGAATTTAGTA
AACCTAGATTTAAAAATGTGCTATAACATAATATATTGAATCTATAATAGTACACCTTGA
CTGCTAAAATATTTCTATAAAATTAATTTGACTTTCCTGATAGAGTTATTCACATCTTATT
TCAACTCACTATAGAAGGAGGAATAGGAGGATTCCTCAGACATCCGGGCATCAGCCCACT
AATGATTTGATTGCTAAGAAAATATTCAGCAATCCAGAAATCACTTGTCAATTTATTCGC
GATATGCTGGACTTGCCAGCAAAAAATGTTGACCATTTTGGAGGGAAGCGATATTCACGT
ATTACTCTCCATGCCTTACTCAGTGCAGGATTTTATAACCAGTATAGACGTCTTGGCGGA
GTTGGATAACGGTACTCAAGTAATTATTGAGATTCAAGTCCATCATCAGAATTTTTCATC
AATCACTTGTGGACTTACCTGTGCAGTCAGGTTAATCAAATCTTGAAAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 383 | 526 | F | 48 aa |

> 3049862-1 ORF translation from 383-526, direction F
VQDFYTSIDVLAELDNGTQVIIEIQVHHQNFSSITCGLTCAVRLIKS*

Description:
unknown

Assembly ID: 3112810
Assembly Length: 885bp

> 3112810 Strep Assembly -- Assembly id#3112810
CTCATCATCTGTCAAAAAGCGTTTCTTAGCAGTCGTGATATCCATAAAATAATCTAATAT
CACGATTTCTCATCCGCAAAGAAAGGAAGGCTGACCAACTCCAGTGCCACATCCTTGTA
AACTACTTCTTGCATATCAAAGTAGGCAAAGTTGAGGTCAGCAGAATCATACCCAATCTG
TTTCAACACTTGACTCTTCATCACTTCAAACCTGACCCTGATCTGTCCCTGTAAATAGGCG
CAGGCTCGGTAAATTCGATAAAGTCAACTTCTGACTTTCTTCAATGGCTAGCATCGTCTC
TCCTTTCTTCAGATTTTTCGATTTAATTTAGTCAATATAGCGCAATTTCCCACGGAAATC
TTCTAAGCTCTCGTAGCCTTTTCCACCATGATTGCTTTCAGTTCATTGGTAAAGCGGTC
AAAAGCACTGACGCCTTCTTTGTGAAGGGTCGTTCCACCTGCACCATACTTGCTCCACA
GAGGATGTGTTCAAAGGCATCTCGACCAGTCAGAACGCCACCTGTTCCGATAATTTGGAT
TTGAGGATTTAAACGTTGATAAAAGGCGTGAACATTGGCTAGAGCAGTCGGTTTGATGTA

TTATCCACCAATTCCACCAAAACCATTTCTTAGGCCGAATAACGACAGATTCGTCTTCTAT
 ATAGAGGCCGTTTCCGATAGAGTTAACGCAGTTGACAACTTGAGCGGATATTTGTTGAA
 AATAGCTGCCGCTTGATCAAAGTGAACAATATCAAATAAGGTGGCAATTTAATTCCAAG
 AGGTTTGGTGAAGTAAGCAAACACTTCTGCCAAAATCCGGTCTGTTGTCTCAAATCATA
 GGCAATCTGAGGTTTACCTGGAACATTTGGACAGGAAAGATTTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 601 | 804 | R | 68 aa |

> 3112810-2 ORF translation from 601-804, direction R
 VFAYFTKPLGIKLPPYFDIVHFDQAAAI FNKYPLK FVNCVNSIGNGLYIEDES VVIRPKN
 GFGGIGG*

Description:

LLCPYRDA NCBI gi: 511014 - Lactococcus lactis. DIHYDROOROTATE
 DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)

Assembly ID: 3112866

Assembly Length: 925bp

> 3112866 Strep Assembly -- Assembly id#3112866

TCTTGGCCAACTGCATGGAGTTCAGCGGTCAATTTCAACGCACCTGAGAAACAGACCCCT
 GCACCCCTGAAATCTCAGGAGACATGATGGTCTGGATGGAATCAATAATGAGAAAGTCTG
 GCTGGATACGCTACCACTTCTGCACGAACACTCTGCATATTTGGTCTCTGCATAGAGATAA
 AACTCACTATCAAATCACCTAAGCGCTCTGCACGTAGTTTAATCTGCTGGGCAGACTCC
 TCCCCACTGACATAGAGAACTGTCCCCACTTGGGACAACCTGGGTTGAGACTTGTAGGAGA
 AGAGTTGATTTCCCAATCCCAGGATCCCCACCGATGAGGACGAGACTTTCCTGGTACAAC
 TCCGCCTCCAAGCACACGGTTGAATTCCTCCATCTCCGTCTTGGTTTCGATTGACATTGAT
 GGAAGTCACCTCAGCTAGTTTCATGGGCTTGGTTTTCTCACCTGTCAAGGACACACGCGC
 ATTCTTGACCTCGGCAACCTCAACCTCTTCCACAAAAGAAGACCAAGACCCACAGTTGGG
 GCAACGTCCCAGATATTTAGGGGAATTATACCCACAATTTTGACATACAAATGTCGCTTT
 TTTCTTTGCGATGACAAACCTCTTTCTATATCTCTAACTCACACTCAATCACTTGGCAAA
 AATCAATCTTCTCATTTGGCACAACCTGGCGCATGAGCATTCGATGAGCAACAACCTACCA
 CAGTCTGATGTTCTCGATACTTAGACATACATTCTAGAAACCGAGACTTCATTTCCGTAG
 CTGTCTCATATTGAATAGGACTATTAGGAAGCAACTCCCCCTTGTTTTCTAAAAACAGTC
 TTCTAGCTGTTTCAAAGTTTTCTATTCTGTCTTTATAGACCTGCCATTCATGTAATAAAG
 GCTCTACTCTTAAAGGAAGACCCGT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| | | | | |

1 220 513 R 98 aa

> 3112866-2 ORF translation from 220-513, direction R
VEEVEVAEVKNARVSLTGEKTKPMKLAEVTSINVNRKTEMEEFNRVLGGGVVPGKSRPH
RWGSWDWEINSSPTSLNPVVPSPGDSSSLCQWGGVCPAD*

Description:

SMS PROTEIN. - ESCHERICHIA COLI.

Assembly ID: 3113664

Assembly Length: 602bp

> 3113664 Strep Assembly -- Assembly id#3113664
TTATGTCAGTGGGATTACGCCTAATCTCCCAGAAGCAGAATTATTATCCGGTCAGGAAAT
TAAACCTTGGNAGACATGAAACTGCAGCGCAGAAATTGCATGATTTAGGAGCGCCAGC
AGTCATTATCAAAGGGAGGCAATCGTCTTAGTCAGGACAAGGCTGTGGATGTCTTTTATG
ATGGACAGACCTTTACTATCCTAGAAAATCCAGTTATCCAAGGCCAAAATGCTGGTGCAG
GTTGTACCTTTGCCTCTAGCATTGCCAGTCACTTGGTTAAAGGTGATAAACTTTTGCCAG
CAGTAGAAAGCTCTAAGGCTTTCGTTTATCGTGCTATTGCACAAGCAGATCAGTATGGAG
TAAGACAATATGAAGCAAACAAAAACAATAAAATCGCCCTTGTATCCCTATTAACCGCC
CTTCTGTGGTTCTAGGTTATTTCTTAAAAATCCCAACACCTACAGGNATTCTAACTCTT
TTAGATGCTGGTGTCTTCTTTGCGGCCTTTTACTTTGGTAGTCGTGAAGGAGCGGTAGTC
GGAGGACTAGCAAGTTTCTTGCTTGACCTCTTATCAGGCTACCCTCAGTGGATGTTTTTT
AG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 165 | 392 | F | 76 aa |

> 3113664-1 ORF translation from 165-392, direction F
VDVFYDQTFITILENPVIQGNAGAGCTFASSIASHLVKGDKLLPAVESSKAFVYRAIAQ
ADQYGVRRQYEANKNN*

Description:

Thi protein - Rhizobium meliloti

Assembly ID: 3113716

Assembly Length: 456bp

> 3113716 Strep Assembly -- Assembly id#3113716
CTGGATACTAAGAGAAATCAAAAAGCACTCTAGGATAGAGGCCTAAAGTGCTTAGTTTC
AAGGCTTTACAGCCTATCATATTTAATAAAATATTACAACATCTTGTTGTAGAATTCAAC

GACAAGTGCTTCGTTGATTTCTGGGTTGATTTTCGTCGCGTTCTGGCAAGCGAGTCAATGA
 ACCTTCCAATTTTTTCAGCGTCGAATGATACGAATGCTGGACGTCCAAGAGTAGCTTCTAC
 TGCTTCAAGGATTGCTGGAACTTTCAATGATTTTTTACGAACTGAGATCACTTGACCTGC
 AGTTACGCGGTATGATGGGATATCAACGCGTTTCCCGTCAACAAGGATGTGACCGCTGGT
 TTACAAATTGGACCAAACCTGACGACCAGTAGTCGCGAGACCAAGACGGTAAACAACGTT
 ATCCAAACGACGTTCCAAAAGAAGCATAAAGTTGAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 94 | 291 | R | 66 aa |

> 3113716-1 ORF translation from 94-291, direction R
 VISVREKSLKVPAILAVEATLGRPAFVSFDAEKLEGLSLTRLPERDEINPEINEALVVEF
 YNKML*

Description:

30S RIBOSOMAL PROTEIN S4 (BS4). - BACILLUS SUBTILIS.

Assembly ID: 3174176

Assembly Length: 1961bp

> 3174176 Strep Assembly -- Assembly id#3174176

CTAATATAGAATAATCACCGCCGTTGTGAAAGAACGATTGGATGATAATCCAATCGTTCA
 GGGAAATTGGAAGACCTTGGGTTTCCAATTTAGGCATGAGACACCTTTGGTGGCTGCTGC
 CGTCCCTCACAAGCTAAGGTGATTGTTGAAAAAGAGGAAAAAGGAGAAGAAATGAAACCA
 GTAATTTCCATCATCATGGGCTCAAATCCGACTGGGCAACCATGCAAAAAACAGCAGAA
 GTCCTAGACCGCTTCGGTGTAGCCTACGAAAAGAAAGTTGTTTCCGCACACCGTACACCA
 GACCTCATGTTCAAACATGCAGAAGAAGCCCGTAGTCGTGGCATCAAGATCATCATCGCA
 GGTGCTGGTGGCGCAGCGCATTTGCCAGGCATGGTAGCTGCCAAAACAACCCTTCCAGTC
 ATTGGTGTGCCAGTCAAGTCTCGTGTCTTAGTGGAGTGGATTCACTCTATTCTATCGTT
 CAGATGCCGGGTGGGGTGCCGTGTTGCGACCATGGCTATCGGTGAACTCTTTTTTAGGATA
 TAAAACAGGGTTCGGATAAGTTTTTTTGAAGGTGGATGATGGCTACATTGTAATGTTTT
 CCTTGTTCTAACTTAGTCTTAAAAGCAGGTGAAAAGTGAGGGCATGCTTTGGCAGCTTGT
 ATGAGTACCTACCGCAGATAAGGGGAACCCCGTTTGACCATCCTCCCAGCTAAATCAATC
 TGACCTGACTGATAAATAGAAGAATCCAGTCCAGCGAAAGCTTGTAATTGAGCAGGATTA
 TCAAAGGCATGAATATTTTGAATCTCGGCTAAAATGACCGCCCCTAAACGATTCTCAATC
 CCAGTAACCGTCGTGATGACCGAGTTTAACTCAGCCATCAAGTCATTGACACATTTTTCC
 GCCTTGTCATGAGCCTCTTGTAATGTTTGATGTTTTTCATTACACGAGATAAACGCTTA
 TCGGTTATCAAACCTATTACCAATTAATAACAAATGTGGTGTAGATCCTTTCCGAAATTGTC
 AAGCGATTGGAGGAAATGAACTAATCCACAGCGGCTTATCCAAGTATACCACTTGGGCT
 TTGGCAGTAGCTAACTGCGCTAAATATAATATAAGGAGGAGTAAAATGAAGACAGTTCAA
 TTTTTTTGGCATTATTTTAAGGTCTACAAGTTCTCATTTGTAGTTGTCATCCTGATGATT

GTTCTGGCGACTTTTGCCCAAGCCCTCTTTCCAGTCTTTTCTGGACAAGCGGTGACGCAG
 CTAGCCAAATTTAGTTCAAGCTTATCAAATGGGCAATCCAGAACTTGTATGGCAAAGCCT
 ATCAGGAATTCATGGTCAATCTTGGCCTGCTGGTTTTGGGTTCTATTTATCTCTAGGTGT
 AATATAAACATGTGTCTCATGACGCGCGTGATTGCAGAATCGACCAACGAGATGCGCAAA
 GGTCTCTTTGGTAAGCTTGCTCAGTTGACGGTTTCTTTCTTTGACCGTCGACAAGATGGC
 GATATCCTGTCTCATTTTACCAGTGATTTGGATAATATCCTCCAAGCCTTTAACGAAAGC
 TTGATTCAGGTCATGAGCAATATTGTTTTATACATTGGTCTGATTCTTGTCATGTTTTCG
 AGAAATGTGACGCTGGCTCTCATCACCATTGCCAGCACCCATTGGCTTTCCTTATGCTG
 ATTTTCATCGTGAAAATGGCACGTAAATACACCAACCTCCAGCAGAAAGAGGTAGGGAAG
 CTCAACGCCTATATGGATGAGAGCATCTCAGGCCAAAAGCCGTGATTGTGCTAGGAATT
 CAAGAGGATATGATGGCAGGATTTCTTGAACAAAATGAGCGCGTGCGCAAGGCAACCTTT
 AAAGGAAGAATGTTCTCAGGAATTCTTTTCCCTGTCATGAATGGGATGAGCCTGATTAAT
 ACAGCCATCGTCATCTTTGCTGGTTCGGCTGTACTTTTGAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 139 | 543 | F | 135 aa |

> 3174176-1 ORF translation from 139-543, direction F
 VIVEKEEKGEEMKPVISIIMGSKSDWATMQKTAEVLDREFGVAYEKKVSAHRTPDLMFKH
 AEEARSRIKIIIIAGAGGAAHLPGMVAAKTTLPIGVVPVKSRLSGVDSLIVQMPGGV
 PVATMAIGELFFRI*

Description:

PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE CATALYTIC SUBUNIT (EC
 4.1.1.21) (AIR C ARBOXYLASE) (AIRC). - BACILLUS SUBTILIS.

Assembly ID: 3174186

Assembly Length: 375bp

> 3174186 Strep Assembly -- Assembly id#3174186
 CTATCTCCAAGTNCGNTTGGGAATNCCTCCGCNANCCACAACCTCATCCAAGCACTTTNCAA
 CGTGNCCTGGTCCGGTCTCAGTGCGTCTNACNGCACCTTCAACCTGCNCATGGGTAGG
 TCACATGGCTTCGGGTCTACGTCATGATACTAAGGCGCCCTATTTCAGACTCGGNTNCCCT
 AGGGCTCCGTCTCTTCAACTTAACCACGCAACAGAACGTNACCCGCCGGTTCATTCTACA
 AAAGGCAGNCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCACACNGCTTCAGGTN
 CTATTTTACCCCCCTCCCGGGGAGCANCTCAACTGACCCNCACGGCACCGGTGNANNAAA
 CGGTCACTTAGGGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-------|-----------|--------|
| ----- | ----- | ----- | ----- | ----- |

1 83 283 F 67 aa

> 3174186-1 ORF translation from 83-283, direction F
VRXXAPSTCXWVGHHMASGLRHDTKAPYSDSXXLGLRLFNLTQQNXTRRFILQKAXSHPL
TGSNLL*

Description:
unknown

Assembly ID: 3174374
Assembly Length: 665bp

> 3174374 Strep Assembly -- Assembly id#3174374
GGGGGGGGTNNNTTCTGGGGCCGGGTGNNTCCTNGAAAAAATGCTGGACTTAACGGTTAA
ATCATTTGAATTGGCCTGTGGATTTTAGCTAGCAATCCAGAGCGAGTTTCTCCAAGACA
GACCTCTATGAAAAGATCTGGAAAGAANACTACGTGGATGACACCAATACCTTGAATGTG
CATATCCATGCTCTTCGACAGGAGCTGGCAAAATATAGTAGTGACCAAACGCCCACTATT
AAGACAGTTTGGGGGTGGGATATAAGATAGAGAAACCGAGAGGACAAACATGAAACTAA
AAAGTTATATTTTGGTTGGATATATTATTTCAACCTCTTAACCATTTTGGTTGTTTTTT
GGGCTGTTCAAAAAATGCTGATTGCGAAAGGCGAGATTTACTTTTTGCTTGGGATGACCA
TCGTTGCCAGCCTTGTCGGTGCTGGGATTAGTCTCTTTCTCCTATTGCCAGTCTTTACGT
CGTTGGGCAAACCTCAAGGAGCATGCCAAGCGGGTAGCGGCCAAGGATTTCCCTCCAATTT
GGANGTTCAAGGTCCCTGTTAAATTTCCCCCATTTAGGGGCAACCTTTTAATGAAANTTT
CCNTNATTTGCCGGGTANCTTTGAATCCCTNGGAAAAAACCCAACNAAAAAAGGGCTTA
NNCCC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 154 | 294 | F | 47 aa |

> 3174374-1 ORF translation from 154-294, direction F
VDDTNTLNVHIHALRQELAKYSSDQTPTIKTVWGLGYKIEKPRGQT*

Description:
REGULATORY PROTEIN VANR. - ENTEROCOCCUS FAECIUM (STREPTOCOCCUS
FAECIUM).

Assembly ID: 3174972
Assembly Length: 989bp

> 3174972 Strep Assembly -- Assembly id#3174972
CTACGATATCTTTGGTCTTTTGTAAGATATGAGGTCCACCCTTATGCGCCTCAGTTGGCA

TTTCATGCGATTCAAGAAGTTGCCCCCTCTTGATCAACCAAACCATACTTGATGTTGGTTC
 CACCGATATCAATTGCAACGTAATATGTCATAAATACCTCCTTTTAGATTAGAGGAAGCG
 CTCCTTGGTTTCACGAATCAAGGCAGCAGCCGCTTCTACAACCTGGACGATCTTCTTCAGT
 CACTGGTGTCAATGGTGAACGAACAGATCCAATATTCAAGCCTTCATTGATTTTCAAGAC
 TTCTTTGATGACACCGTACATATTTCCATGAGCAGAAGTGAGTTTACCAATGATTGCGTT
 GATAGCATACTGCAATTCACGCGCTGTTTCTAGGTCCTTATCCGCAATCAACTGATTGAG
 TTTCAAGAAGAGTTCTGGCATAGCACCATAAGTACCACCGATACCAGCCCCTAGCCCCCAT
 GAGGCGTCCTCCTAGGAAGTGCATCAGGACCATTAAAGACGATATGGTCTTCTCCACC
 AAGGCTGACAAAGGTTTGGATATCTTGAACCTGGCATAGAAGAGTTCTTCACACCGATAAC
 ACGAGGATTTTTTCAACATTTCTGTGTAAAGGCTTGGAGTCAAAGCAACCCCTGCCAATTG
 AGGAATGTTGTAAATCACGTAGTCTGTGTTTGGAGCTGCAGAACTGATATCGTTCCAGTA
 TTTGGCAACTGAGTTATTCTGGCAAGCGGAAATAAATTGGTGGAATCCGTTGCAATAGCA
 TCTACTCCCAAGCTTTCAGCATGGCGAGCAAGTTCCATACTATCTTTAGTATTATTGCAA
 GCAACATGGGCAATAATGGTCAATTTACCTTTGGCTACCGCCATGACTTCTTCCAAAATC
 AACTTGCGATCTTCAACGCTTTGGTAGATACATTCACCAGAAGAACCATTGACATAAGAC
 CTTGAACACCTTTTATCAATGAAGTATTGA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 169 | 678 | R | 170 aa |

> 3174972-1 ORF translation from 169-678, direction R
 VIYNIPQLAGVALTPSLYTEMLKNPRVIGVKNSSMPVQDIQTFVSLGGEDHIVFNPGDEQ
 FLGGRLMGARAGIGGTYGAMPELFLKLNQLIADKDLETARELQYAINAIIGKLTSAHGNN
 YGVIKEVLKINEGLNIGSVRSPLTPVTEEDRPVVEAAAALIRETKERFL*

Description:

N-ACETYLNEURAMINATE LYASE SUBUNIT (EC 4.1.3.3) (N-
 ACETYLNEURAMINIC ACID ALDOLASE) (N-ACETYLNEURAMINATE PYRUVATE
 LYASE) (NALASE). - ESCHERICHIA COLI.

Assembly ID: 3175138

Assembly Length: 1450bp

> 3175138 Strep Assembly -- Assembly id#3175138

CTCCATATTTCTTAGCCTTCTCAATTAGGGTCTTGAAGTCTTCGACACCACCGATACGCT
 TACCAATATCAGCATAGTTCAAGTGACCAGAGTCATGGCTGTGATATCCTTAACTTTTTC
 CCAACCTTGAGGGTTGTTTATAATGCTACGATAAGCAATGGCACCATCTTGCCAATCAAC
 TTTCTTGTCTGCATTGGCATCTTCAGTGATAACAACCTTAGCACTTGGAAAGTTCCTTCGT
 GTATTCTGGGAAAACAATGCCCTTATAAGCTTTTTCCCATTTGCCATTGAGAGCTGTGGAT
 TCCTACATAGTTGGCATTTCCGACTGTTTCTTTATAAGCTGTCAAACGAGTCCAGTCATT
 CGAACCACCACCATAGCTATTTTGAGAGTTACTCCAAACACCAGCAGCAAGCTTATCTGT

AGAAACAAATCCATACATGTAACCCCTTAGCCAAATCCTTCATTGGATTGGTTACATCGAT
 ATGATCATCTCCGCTGACATGCGTATTGTTTGACATGGTTGCCCCATCAAACCTTAGCACC
 AGTTTGATCACTAGAAACAGAGACTAAAGCATTGCCGAGGAACTAATAGAAGAAAGTAG
 TTTTCTTTTCGTCATCAATCTTTTGACCTGGAGTGACTTGATTGTGGTTGACAATCTTGGT
 CACATCAAAGTGCAATTGATTGTCCACAACCTTGCAAGCGTACTGTCATTTCCGCATTGAT
 TAAGTGAGCATCATCGCGAAGCTTCATCAAGTACTCTGCTGTTGTCTCATTGATTTTTTT
 ATAAGTGACTTCAGGGGTGATTCCGGTGGTTATTGATAAAGACTTGGGTTGAATTGTTGCAC
 CTGTCCTGGCAAAGTATGTCCATTCAAGGTGTATCCCTTGACACGAAGGAAGGCTTGGTC
 AATTACTGCCTTAAGTACCTTAAACTGGATCGTATCATAAGTCACCTTGCTATCGTCAAC
 AACCGGACCTGTTTCTTTCTGGGCAGGGGTATCCTCTGGGTTTTACCCTCTCTGTGGCTA
 TCCGTTTCAACGCTTGAACAACTGGTCGCTCATCGTCATAAGAGCCCGCCTTGAGAAAAA
 TCTTCTTCTCATTTCTAAGATGGTCATTGACCGCAGCTGGTAGAGTCACTGTGTCAAAGA
 AGATTGACATCCTTATTTGCCTGGCATTACCTGACCGTCTGACTTGAAGACTGATAGAG
 AGACGGTTTGTGATCCTGTTTCAGGAGCAGCAACACGACTACCTCTATACCAAGTGCTA
 GTTGTGGAGATTTATACTCCCAGAACCAGCCATCCTTGTCATAACCGACAAAAACATTA
 TTATTGGTATCTTTAAATTTCAAGGAGACACCAAAGCGTGATTTGCCCTTTTCAGAATCT
 TCTTTGAAGGTTAAATCAACAGTTGCATTTCCATTGGCATCAACGGTCAAGCCCTTCTTT
 TCAAACAGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 79 | 945 | R | 289 aa |

> 3175138-1 ORF translation from 79-945, direction R

VTYDTIQFKVLKAVIDQAFLRVKGYTLNGHTLPGQVQQFNQVFNNHRITPEVYKKINE
 TTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIVNHNQVTPGQKIDDERKLLSSI
 SFLGNALVSVSSDQTGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMYGFVSTDKLA
 AGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELP
 SAKVVITEDANADKKVDWQDGAIAYSIMNPNQGWKVKDITAMTLVT*

Description:

unknown

Assembly ID: 3175860

Assembly Length: 420bp

> 3175860 Strep Assembly -- Assembly id#3175860

CTGCGAGTTGTGAGGCTCCTATTATGTCTCGTGATTAATCTCTATAAGGTGATTTTGG
 AGGGAAATTATCGGGCGACAGCGGGTAGAGAAGAGATGAAAGAGGCTATTTTGGAAATATC
 AAGCAAATCCTGCTGCCTTAAAGATCTCAAAGAAAAGGCTAAGAATATTTCCAGAGAGT
 ATTCTGAAGAGCATCTGTTACAAATCTGGTTGGACTTTTATGAGAAACAAGCCGCTTTAG
 GGACAAAGTAAAAAGTGAGGTAATCTATGCGAATTGGTTTATTTACAGATACCTATTTTC

CTCAGGTTTCTGGTGTGCGACCAATATCCCAACCTTGAAAACCCACCTTGAAAACACGG
 ACTTGCCTGCATTTNTATCTCATACAATCCACCGAATTTTCGATGTCCCCCTCCCTACAAC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 51 | 251 | F | 67 aa |

> 3175860-1 ORF translation from 51-251, direction F
 VILEGNYRATAGREEMKEAILEYQANPAALKDLKEKAKNISREYSEEHLQLWLDIFYEQ
 AALGTK*

Description:

unknown

Assembly ID: 3175918

Assembly Length: 661bp

> 3175918 Strep Assembly -- Assembly id#3175918

CTCCCCAACTTTTATTTGAGAGTGAACGGTATAAGAATATGAAACCGGAGGTTAAGGTG
 GTTTACTCAGTTTAAAGATCGGTTGGAGTTGTCTTTGAGCAAAGGTTGGATTGATGAG
 GATGGGACTATTTATTTGATTTATTCCAATTCAAATTTGATGGCACTTTTAGGCTGTTCA
 AAGTCAAAATTACTCTCCATGTGAGTTGAAGTGACATTTTTAGATGATTACCATAAAAA
 ACATAACTACCCACTATTTTACGAATCCTATCTTCAAAACGTTATGGAATTCCTTGAAAG
 TCAAGACATAAAGAATGGGGTTGATGCCTTTGTAGATGATCATCAAAATCTCGTTTTTGT
 TTTATATGGACAAGGCTATCGAGCCGAGGGAAAAGAGGGAATACTTACAACCCAAGTAAC
 TGTAAGGCTTATGATGAAGACAAGAAACCGATTAACTTCGCAAAATTTATTAGATTCCTT
 AATCGTGTGAGAATATCAAATGGAACCGAATCTTTGGGAGGTCTCCTATGATTGATCTCT
 ATCTAAGTAAAAATAGCCGAAGAAATCAACTTCTTTTAGACTTCTTCCAAACTATGGCA
 TCGAGGTATCTTGTCAATTCAGTTTCTGAAATGACAAAGGACAAATTAATTGAGATGATGA
 G

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 212 | 535 | F | 108 aa |

> 3175918-1 ORF translation from 212-535, direction F
 VTFLDDYHKKHNYPLFYESYLQNVMEFLESQDIKNGVDAFVDDHQNLVFLYGGQYRAEG
 KEGILTTQVTVKAYDEDEKKPINFANLLDSLIVSEYQMEPNLWEVSYD*

Description:

unknown

Assembly ID: 3811220
 Assembly Length: 1429bp

```
> 3811220 Strep Assembly -- Assembly id#3811220
CTGCCCCCTGTAAGGCTGGACGATTGCCTTTCTTAGTATCCGCAAAGAGGTAAACTGAGAA
TAGAGAGGATTTCTCCTTCAATATCTTTGACAGACAGGTTTCATCTTGCCTTCTACGTCTG
AAAAAATCCGCATATTGACCAGTTTCTCACAGCATAGTCCAAATCTTCCTCTTGGTCCT
CTGGTCCAACACCAACCAGCAATAAAAGTCCCTGATTGATTTTTCCCTGAATCTGGCCTT
CTATACTCACTTGGGCTTTTTTAACCCGTTGGATAATGATTTTCATAATAGCCTTTCTAG
TAAGAGCTAGGACAACCTAGCCGTTGGTCCGTTTGACAGAGTAACTTCTGGCACACTCTT
AATTTTATCGACAACCGTGGTCAGTGTAGAGAGGTTGGCAATACCGAAGGACACATGGAT
ATTAGCAAACCTTCATATCCTTGGTTGGTTGGGCATTGACCGTTGAAATATTCTTGGTTGT
ATTTGAAAGAACTTGCAGTACATCGTTCAACAGTCCTGTACGGTTGAGACCGTAGATATC
GATATGGGCCATATACTCCTTATTTGAGCTAGAGTACTGGTCTTCCCATTCCACATCAAG
GAGACGTTGCTCGTAGTTTTCTTGGGCACGCAGGTTTCATACAGTCCACACGGTGAATAGC
CACACCACGACCCTTGGTAATGTAGCCAACAATATCGTCAACCAGGCACGGGGTTACAACA
CTTAGCAATCCGCACTAGGAGACCAGAAGCACCTTCAATAACCACTCCCCCTCATGCTT
GACCTTGGAGAGTTTCTTTATTTTCAACCTTGACCTCGCCACCTTTGACAAGCTCCTCTG
CCTCAGCCTTGGCCTTGGCACGCTCTTCTCACGGCGTTCTTTTTTCAGTCAGACGGTTAA
AGACGGTAATCGCACCGATTTCCCCAAAACCAATGGCCGCAAAGAGGGAGTCTTCTGTCT
TGTAACGGTCTTTTGCAGAACTTGATCCATGTGGCGCTTGTCCATAAATTTATTTGCCA
CATAGCCATTTTCTTGGAACTGAGCCATCAGCATCTCACGACCCTTGTGACAGACAATT
CCTTATCTTGGTTTTTAAAGAACTGGCGAATCTTATTGCGCGCCTTGCTAGTCTTGACCA
TATTGAGCCAGTCACGGCTAGGTCCAAAGGAGTTCGGGTGGCGATAATTTCAACCTGAT
CCCCGTCTTTAACTTGGTTGTCAGTGGAACCATGCGGCCATTGACCTTGGCACCAAGTTG
CTTTTTCACCGACCTTGGTATGGATTTTCGTAGGCAAAATCAATCGGTCTCGAATCTTTGG
GAAGAGAACGGACAGCTCCATCTGGGGTAAAAACGTAAATCTCCTCAGCCAGATAGTTTT
CCTTAACAGAGTCCACAAATTCCTTAGCATCATCAGCCTGGTCTTGGAG
```

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 316 | 873 | R | 186 aa |

```
> 3811220-2 ORF translation from 316-873, direction R
VRKSVPRPRLRQRLSKVARSRLKIKKLSKVKEGGVVIEGASGLLVRIAKCCNPVPGDD
IVGYITKGRGVAIHRVDCMNLRAQENYEQRLLDVEWEDQYSSSNKEYMAHIDIYGLNRTG
LLNDVLQVLSNTTKNISTVNAQPTKDMKFANIHVSFGIANLSTLTTVVDKIKSVPEVYSV
KRTNG*
```

Description:

stringent response-like protein - *Streptococcus equisimilis*

Assembly ID: 3811436
 Assembly Length: 1513bp

> 3811436 Strep Assembly -- Assembly id#3811436

CTCTGCAATGATGTACTCAAACATCTCCGCTTCTAGTTCCTCCTTAGGCAGAGGCAATTT
 CCCACGTCGCATCCGGTTCATAAAGACCGTATGGTTTTCTAAAATCAAACATATACAAAC
 CATGTGGGGAATATCCAATCCAATGGCTTTAGCCACATTTTCCTTTACTTGCTCCATGGT
 CTGACCAGGCAGAGCATAAATCAAATCAATGGAGATGTTGTCAAACCAGCCAGTTTCAG
 GCGATCGATATTTTCATAAATATCCTTCTCCAAATGACTGCGCCCAATCTTTTTCAACAT
 CTTATCATCAAAGGTCTGGACACCTAGCGAAACACGATTGACAGCCGAATTTTTCAAAAC
 AGCTATCTTATCCGCATCCAAATCGCCTGGATTGGCTTCAATGGTCAACTCTTCCAAGAC
 AGACAAATCCAAGTTTTTAGTCAAGCCATTCAAGTAACACCTCCAGTTGCGGAGCCGACAG
 GGCTGTCGGTGTTCACCACCGATATAAAGGGTTGACAACTTTTCAATATCATAAGAACG
 AAACCTCTTCCAGCAGATGCTCTAAATAGCTGTCGACTGGCTGATTTTTGATGAAGACCTT
 TGAAAAATCACAATAATAACAAATCTGGGTACAAATGGGATGTGCACATAGGCTGACGT
 TGGTTTTTTCTGCATAGTAATTATTATACCACAAAGACTAGATTCCAGATAAAAAATCACC
 ATCCCCAGATACATAGTCCGTCCGGAGATGGTGATGGTTTATTCTTCTGTTATATCAATC
 ACAATCTCTTCTGAGTCATCAAGAGCTTCGGCTTTTTTCTTGCCATTGTTTCCTTGAGATTA
 TTTAATTGATTTTTTTGATGCTTCTGTGCTTGAAAAGCATAGGATTTAGCTTGAGCAAGT
 ATACTGTCCACAGTGATTTACCTGACTCAACCTGTTCTTTTGTGTTTTCAGAACAAAATCT
 GTAGCCTGCTCCTTAACCTCTGTCAGTTTTTCACAGACTTGCTCCTTGGCATACTCCGGA
 TCTTCTCTCAAATCATCTAAAAAATCTTGAGCCTGACTGCAAACCTGTTTGCCCTTATCA
 CTTGTGTTAAAAACAAGGCAAGAGCTGCACCTGAAACGGTTCCTAAAAGGATTGAGGATAAT
 TTACCCATAAGGATTCTCCTTTTTTATTTTTTGAAAAATTTACTTGCAAGACGAAGAGCT
 GACAGACTTGACACCAGTCTTGAGTGTTTTTGAACCAGCTGATGAAGCTTTCTTGCTCAAG
 ACACGCGCATGGTCATTGAGGTCTGAAACAGATAGAGATAAATCTGCAACAGCACTGAAG
 AGTGGATCAATCGTAGCCACCTTGACATTGATATCATCTGCCAAGACATTGACCTTAGCC
 AACAACTCATTGGTGTGATGCAAGGTCACATCCACATCTGAAGTCAAGGTTTTAATCGTC
 TTTTCTGTTTCATCGATGACACGACCAAGCTTTTGTACAGTAATGATCAGATAGACCAA
 AAGACAATCACAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1164 | 1511 | R | 116 aa |

> 3811436-3 ORF translation from 1164-1511, direction R
 VIVFLVYLIITVQKLGRVIDETEKTIKTLTSDVDVTLHHTNELLAKVNVLADDINVKVAT
 IDPLFSAVADLSLSVSDLNDHARVLSKKASSAGSKTLKTGASLSALRLASKFFKK*

Description:

unknown

Assembly ID: 3811984
 Assembly Length: 505bp

> 3811984 Strep Assembly -- Assembly id#3811984

CTCTTGTCTCAGAGAAATTTACAAAACGTTAGGAGAATAAGATGGCATTATTGAAAAAGGT
 CAAGAAATCGATATGGAAGTCATCAAGGCTGAAACCCAATTGTCTGCAGAAGCCTTGAGA
 CTCAAGGAAAGCCGTGACAGGGAATTGGCAGATATTATTTTCAGGGGAAGATGACCGTATT
 CTCTTGGCTGATTGGTCCTTGCTCTTCTGATAATGAAGAGGCGGTCTTGGAATATGCTCG
 CCGTTTATCCGCCTTGCAAAAGAAGGTAGCGGATAAGATTTTCATGGTCATGCGCGTGTA
 TACTGCTAAGCCTCGTACCAATGGAGACGGCTATAAAGGGTTGGTTCACCAGCCAGATAC
 TTCTAAGGCTCCAACCCTGATTAACGGCTTGCAGGCTGTGCGCCAGTTGCACTACCGCGT
 TGATTACAGAGACTGGTTTGACAACGGCAGATGAGATGCTTTATCCGTCAAATCTGATCT
 TGGTGGATGACTTTGGTCACCTACC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 134 | 454 | F | 107 aa |

> 3811984-2 ORF translation from 134-454, direction F

VTGNWQILFQGKMTVFSWLIGPCSSDNEEAVLEYARRLSALQKKVADKIFMVMRVYTAKP
 RTNGDGYKGLVHQPDTSKAPTLLINGLQAVRQLHYRVVDYRDWFDNGR*

Description:

PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE, TYR-SENSITIVE (EC
 4.1.2.15) (PHOSP HO-2-KETO-3-DEOXYHEPTONATE ALDOLASE) (DAHP
 SYNTHETASE) (3-DEOXY-D-ARABINO-HEP TULOSONATE 7-PHOSPHATE
 SYNTHASE). - ESCHERICHIA COLI.

Assembly ID: 3857228
 Assembly Length: 1827bp

> 3857228 Strep Assembly -- Assembly id#3857228

CTCTTTTAACCGTTTTAGCGGTGACACCGAGGATTTTTTCAGGACCCAAGACTTGTCTGGG
 CAACCGAAACTGGGAGTTCGTCATCTCCAATATGCAGACCAGCAGCATCAACCGCAAGAC
 AAACATCCAACCGATCATCGATTATCAAGGGGACCTGATAGGCATCTGTTATTTCTTGA
 CTTGTTTTGCCAGTTGATAATATTGATTGGTTGTGAGATTTTTTCTCGCAATTGGACTA
 TGGTAACCCCTGAACGGCAGGCCGTCTCAACTTTTGCAAGAAAGCTTTCCACGGAATCTT
 GATAGCGATTGGTTACCAGATATAGTCTAAGCGCTTCTCTATTCATAAACCTCTCCTTTG
 ATGGTATCTAGCCAATTTTCATCTCTTCTTAGGAGCGAAAGCTGATTGAGTACTTGGTAA
 CGAAATCTTCCAATCCCATTCTTGAACAATATTTCTCAGCAGCGATATTGAGATAA
 GAGACTGCTAAGCAAGAACTTCAAACACAGTCTTTCCTTGGCTGAGAAAAACAGCTGTTA

AGGCTCCAACCAAGTCTCCTGTCCCTGTTATCCAGTCTAATTCAGTACAGCCATTCTCAA
 GTACAGCAACTTGATTCTCCGAAACAATAAGGTCCCTTGGGACCTGTGACTAAGAATGACA
 TACCACGATAGGTCTGACACCAGTCTTTCAAGACTTGAAGCAAATCCTCCGTTTCTTGAT
 CTTTAGCACTCGCATCGACCCCAACGCCGTGATGCTTTAATCCAACAAGACTTCGAATTT
 CTGACATGTTTCCTTTAAGGACCGTAGGTCTATAGTCTAAAAGGTCTTTAACTAAGCTCT
 TACGAATGGATGAAGTCGTTACGCCAACCGCATCTACTACCATCGGGAGAGAAGATTGGT
 TTGCATACAAAGCTGCCATGCGGATTGCTTTTTTCCTTCTCAGCTGACAAATGCCCCAAAT
 TGATGAAGAGAGCCTGGCTTTGCTTAGTAAAATCAAGAACTTCACGGGGATCATCTGCCA
 TGACAGGTTTGCATCCCAGAGCCAAAATCCCATTTGCCAGCATCTCACAAGAAATCTCAT
 TGGTCATACAGTGAATGAGGGAAGTAGAGCCTATAGGAAAAGGATTTGTCAATGCCTGCA
 TCATTCTATCCTTTTCAAGCAAAGAAATATCCTTGCACTTTTTTAAAGAATTCCTGCTTGAT
 TAAAAATCTAAATGCAATAAAGGAAATCGCTGTACCAATCAAGGTTGCTCCGAAAAATCG
 AGGCGTGTAGATAAAACCAACTAAGCTTAGCAGCCGATCCTGTAAAGAGCACCATAACAGG
 ATAGGAAACAATAGAACCAATAATACCTGTTCCCACAATTTCTCCCAAGGCAGAAAAGTA
 AAATTTTCGACCGTACTTATAAAAGAGACCTGCTAGAAGGGCTCCAAAAGTCGCTCCTGT
 GAGAGATAAAGGAGCTTATCGGAATACCCTTGAGTCGTCATACGGATAAAGGCTGTCACT
 GTAGCCATAGCCAAGGCATAAACAGGTCCCATCATGATTCCCGCTAGAATATTGACTACA
 CTGGACATCGGTGCCATTCCTCAATCCGAAAGATAGGTGTAAGGACTACATCAAGGGCA
 ATCATCATAGATAAAAATGGTCAATTTGTGAACTTGTAGTTGGTGCTTTCTCAAGTTTCTA
 TTCTTCTCCTTTTTTCTAAAGACTGTAAATCGCTCTTCCATGTCTGGTGTGGTAAGCCAT
 CTCCCAAACTTGGCTTCCATATGAACACTGATGTGGAAGGCATCTAGCATTTTTTGCTT
 ATCTGTCTCATCACTTTCTCGATAGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1141 | 1356 | R | 72 aa |

> 3857228-2 ORF translation from 1141-1356, direction R
 VGTGIIGSIVSYPMVLFTGSAKLSWFIYTPRFFGATLIGTAISFIAFRFLIKQEFFKK
 VQGYFFAERIE*

Description:

unknown

Assembly ID: 3857842

Assembly Length: 485bp

> 3857842 Strep Assembly -- Assembly id#3857842

CTATTGCCAATCCATATAGCCTATCAGGTGGTCAATAACAACGTGTGGCCATCGCTCGTG
 GCCTATCAATGAATCCAGACATCATGCTCTTCGATGAACCAAATTCGCCCCTTGACCCTG
 AGATGGTTGGAGAAGTAATTAACGTTATGAAGGAATTGGCTGAGCAAGGCATGACCATGA
 TTATCGTAACCCATGAGATGGGATTTGCCCCCAGGTTGCCAACCGCGTTATCTTTACTG

CAGATGGCGAGTTCCTTGAAGACGGAACACCTGACCAAATCTTTGATAACCCACAACACC
 CTCGTCTGAAAAGAGTTCTTAGATAAGGTCTTAAACGTCTAAACTCAAACCTGCAAGGATTT
 CCTTGCAGTTTTTCTACCTCGTATTGGAATTTTTGATTTTTTCGGAAAATTATGTTAGAAT
 TAAGTTTATGAAATGAGGTTTCCTCATACCTAGCAAGACTAGGAATAAAAATAGAAATTA
 GGTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 45 | 341 | F | 99 aa |

> 3857842-1 ORF translation from 45-341, direction F
 VAIARGLSMNPDI MLFDEPN S ALDPEMVGEVINVMKELAEQGMTMIIVTHEMGFARQVAN
 RVIFTADGEFLEDGTPDQIFDNPQHPR LKEFLDKVLNV*

Description:

GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ. - BACILLUS
 STEAROTHERMOPHILUS.

Assembly ID: 3857996

Assembly Length: 1547bp

> 3857996 Strep Assembly -- Assembly id#3857996
 NTCTTGGGCNCNGGGCGNNTCCTTTGAGGACNACGGTATCGATGACCTTGATCTCAAGTG
 CAAGCAGTATCTGAATCTGCAGCAGCACCTGTCCGTGCAAAAGTTCGTCCAACATACAGT
 ACAAACGCTTCAAGTTATCCAATTGGAGAATGTACATGGGGAGTAAAAACATTGGGCACCT
 TGGGCTGGAGACTACTGGGGTAATGGAGCACAGTGGGCTACAAGTGCAGCAGCAGCAGGT
 TTCCGTACAGGTTCAACACCTCAAGTTGGAGCAATTGCATGTTGGAATGATGGTGGATAT
 GGTCACGTAGCGGTTGTTACAGCTGTTGAATCAACAACACGTATCCAAGTATCAGAATCA
 AATTATGCAGGTAATCGTACAATTGGAAATCACCGTGGATGGTTCAATCCAACAACAACCT
 TCTGAAGGTTTTGTTACATATATTTATGCAGATTAATTTACAGAGGGACTCGAATAGAGC
 CCTCTTTTCAGGTTTTACCGTGACAATCCCTATTAAAAATTATATCAAAATCGTGAAAAT
 ATTGGAAGATATGGTAGAATGAAAATTGTCGTGTGAACGATAATACTCATTCTTGATGA
 ATTGTGAAGCAGTTGCCCTTGGGTCGTTTTGCGAGTTGAAGTCAAGAAGAGGAAAAAAC
 AAAAAGGAGAAATACTCATCGAATTTCAATGAAACAACCTTCTTGAGGCTGGTGTACACTT
 TGGTCACCAAACCTCGTCGCTGGAATCCTAAGATGGCTAAGTACATCTTTACTGAACGTAA
 CGGAATCCACGTTATCGACTTGCAACAAACTGTAAAATACGCTGACCAAGCATACGACTT
 CATGCGTGATGCAGCAGCTAACGATGCAGTTGTATTGTTTCGTTGGTACTAAGAAACAAGC
 AGCTGATGCAGTTGCTGAAGAAGCAGTACGTTTCAGGTCAATACTTCATCAACCACCGTTG
 GTTGGGTGGAACCTTTACAAACTGGGGAACAATCCAAAAACGTATCGCTCGTTTTGAAAGA
 AATTAAACGTATGGAAGAAGATGGAACCTTCGAAGTTCTTCCTAAGAAAGAAGTTGCACT
 TCTTAACAAACAACGTGCGCGTCTTGAAAAATTCTTGGGCGGTATCGAAGATATGCCTCG
 TATCCCAGATGTGATGTACGTAGTTGACCCACATAAAGAGCAAATCGCTGTAAAGAAGC

TAAAAAATTGGGAATCCCAGTTGTAGCGATGGTTGACACCAATACTGATCCAGATGATAT
 CGATGTAATCATCCCAGCTAACGATGACGCTATCCGTGCTGTTAAATTGATCACAGCTAA
 ATTGGCTGACGCTATTATCGAAGGACGTCAAGGTGAGGATGCAGTAGCAGTTGAAGCAGA
 ATTTGCAGCTCCAGAACTCAAGCAGATTCAATTGAAGAAATCGTTGAAGTTGTAGAAGG
 TGACAACGCTTAATTTATACAAATAGTAATTACCTAGGAGGGCGGGGCTTAGCCCGGCTC
 TCCTATTTTCAAAAAATATAGGAGAATTAAAATGGCAGAAATTACAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 58 | 456 | F | 133 aa |

> 3857996-1 ORF translation from 58-456, direction F

VQAVSESAAAPVRAKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWNGAQWATSAAAA
 GFRTGSTPQVGAIACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTT
 TSEGFVTYIYAD*

Description:

unknown

Assembly ID: 3858236

Assembly Length: 740bp

> 3858236 Strep Assembly -- Assembly id#3858236

CTATAAAAAAAGGGTAACCAGTATGGAGGATGAATGTCTGGAATATCTGAGAATCTCG
 GATTTTGGAAATCAGACCGATCATCATGAGATAAGGAAGGAAAGCACTTGTA AAAAGCAC
 TGTAACCACGCCAGTCCCCTGTCCCAAGAGGGTGAGGTGGTAGCGTAAAACCATGCGGAA
 AAATCCCTTTTTTAGTGGTTGAAATTCTCTCCTTGCTGCGACGTTCTTTTTTGACCTTCTC
 CTCACTATTAAGCAGGATCACGTCATAAAAACGAGGAAGGACCTTCTTTTTTGGTCAGATA
 AAGCAGGAAGAGAGTTAGTCCTATCCAAGCGAGCAGACCCAATATGGCTTCTATTGAAAA
 AGGCTCCACTGCTATTTTGTAAAAGATATGAAGAGGATAAAGGAGAAATGGAATGTCTCT
 AACTTTGTCAACAATACTTCCAAAAGTCGACTGAAGAAAGAAGATAAATATTAAAGGTAT
 GAGAACTCCTATCCCAATCATCACATTCGAAAAAATAGACTGATACTTTCTGAAGACCCT
 AGTCTGAGCCAAGAAATGTACTGCCACTACCGTCACTAAAGTAACAGAGACAAATAATAA
 GGTCAAGGACAGTAGCATCAAAGGCAAACCCAGCCAAAGAGAAGGAGCTAGACTAATATA
 GAGGGCTAGAAAATAAGCTAGGATTGGTACAATTCCAGTTAGAGCTGGCAAGAGGACAGA
 CAGTCCTTTAGCAATTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 1 | 261 | R | 87 aa |

> 3858236-1 ORF translation from 1-261, direction R
VILLNSEEKVKKERRSKERISTTKKGFFRMVLRHLTLLGQGTGVVTVLFTSAFLPYLMM
IGLISKIRDSQIVPDIHPPYWLPPFL*

Description:
unknown

Assembly ID: 3858264
Assembly Length: 2219bp

> 3858264 Strep Assembly -- Assembly id#3858264
ATCGAATTCGTTTTGCAAGTGGCGAAATGCGAACCACGTTTGTGTCTTTATAAGTTTCCA
CGTCTTCTTTGTGGACACGACCGTTTGCACCTGAGCCAGAAACGTCGTAGAGGTTTATCC
CTAAATCATCCGCTAACTTTCTAGCTGCAGGAGTCGCTCTTAGCTTGTCATCAGCCATGA
CCTCTCCAATTCTATTTATGATACAAAGGGCGTCAAAAGCGACTGAAAAATAGGAAATCG
ACGATGGCTTCGATGAAGCCAAGGAGATTTATCTTTTTTCCAAGCTTTTAGCCCGTGCT
CTAATCTAAGATATTAAGGACGAAGAGCTCTGCACCTAAAAGATACAAAGTTCTCGTCAG
CTTTGTTTTATTTACATAACTTATCTTATGTAACCTCTATTCTTTGTTATAAGTTTTTCGG
ATTGCATCTTTGATACTTTCAACTGTTGGAATCATTCACATTTTTAGGTTTTGCGCATA
AGGCATCGGCACATCTTCTCCTGCACAACGGCGGATTGGTGCATCTAGATAGTCAAATGC
TTCTGATTCTGAAATAATAGCTGAAATTTACCGATATAGCCACTTGTTTTGTGGGCATC
GTTGACCAGAACAACCTTACCAGTCTTCTTCACTGAGTTTATGATGATATCCTTATCAAG
CGGAACAAGGGTACGTGGGTCAACAATTTCAACTGAAATTCCTTCTTCAGCTAATTCTTC
AGCAGCTTGAACCACACGGCGAAGCATTTTTCCATAAGTGACAACGTGTACATCCGTTCC
TTGGCGTTTGATTTACCAACCCCAAGTGAATTTGTGTAGTCTGGATCAACTGGCACTTC
CCCTTTTTGGTTAAATTCTGACTTGTACTCAAGTATAATAACTGGGTGTTATCACGGAT
AGAAGACTTAAGCAGGCCTTTCATGTCCGCAGGTGTTCCAGGTGCCACAACCTTAAGCCC
TGGAATGTGAGTAAACCAAGACTCTAGAGATTGTGAGTGCTGGGCGGCAGAGCCAACTCC
GTTACCAGCTGCACAACGAACAGTCATTGGAACCTGACCTTTACCACCAAACATGTAACG
TGTTTTAGCAGCTTGGTTGACGATATTGTCCATGGCAATAACAGAGAAGTCCATGAAGGT
CATATCGACGATTGGACGAAGTCCTGTCATGGCTGCTCCTGCTGCAGCTCCAGAGATGGC
AGCTTCAGAAATCGGACAGTCACGGACACGTTCTGGACCAAATCTTCAAGCATTTCCAAC
AGAAGTACCGAAGTCTCCTCCGAAGACACCGACGCTTCTCCCATCAAGAACACATTTTC
ATCGCGAACGCATTTCTCAGACATAGCAAGGATAATGGTGTACGGAAGGACATTGTTT
TTGTTTCCATTTTATCTCTTTCTCCTTAGTCTGCGTAAATATCTTCAAAGGCTGATTCAA
GCGGTGGGAATGGGCTTTCTCTGCAAATTTAACAGAAGCTTCTACTGCTTCCTTTACTT
GCGCTTGGATTTCTTCCAATTCTTCGGCACTTGCAATGTTATTTTCAATAAGGTAATTGC
GGAGGTTTTTCGATTGGATCTTTTTGTTTTCCACAATTCCACTTCTTCACGCTACGATATT
TACCAGGGTCAGATGATGAGTGACCGAGCCAGCGATAAGTTACACTTTCAATCAAGACTG
GACCATTGCCACTGCGAACATGGTCTATAGCTTTCTGAAATCCTTCATAGACATCGATGA
CATTGTTACCGTCTTCGATGAACATTCAGGAATTCATAAGCGGCGCTACGTTGATGGA
TATGTTCTATATTGGTCATTTTCTTGATATCCGCAGAGATACCGTAACCGTTGTTAATGC
AATAGAAAATGACTGGCAGGTTCAGATAGAAGCCATGTTCACTGCTTCGTGGAAAACAC

CTTCATTGGTCGCACCATCTCCAAAGAAGCAGACAACGATTTTACCGGTATTTTGCATTT
 GCTGACTGAGGGCTGCACCGACAGCGATCCCCATACCACCACCTACGATACCATTTGGCAC
 CAAGGTTCCCAGCATCAAGGTCAGCGATATGCATAGATCCACCTTTCCCTTTACAGGTTT
 CAGTGTATTTACCAAGGATTTTACGCCATCATTCCGTTGAAGTCAATCCCTTTAGCAATAG
 CTTGCCCCGTGTCCACGGTGGTTTGAGGTAATCAGATCATCTGGATTGAGAGCTACATAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 439 | 1365 | R | 309 aa |

> 3858264-1 ORF translation from 439-1365, direction R
 VTPLSLCLRKCVRDENVFLMGEDVGVFGGDFGTSVGMLEEFGERVRCPISEAAISGA
 AAGAAMTGLRPIVDMTFMDFSVIAMDNIVNQAAKTRYMFGGKGQVPMTVRCAAGNGVGSA
 AQHSQSLESWFTHIPGLKVVPAGTPADMKGLLKSSIRDNNPVIILEYKSEFNQKGEVPVD
 PDYTIPLGVGEIKRQGTDTVTVTYGKMLRRVVQAAEELAEEGISVEIVDPRTLVPPLDKDI
 IINSVKKTGKVVLVNDAAKTSYIGEISAIISESEAFDYLDAPIRRCAGEDVPMPYAQNL
 KMCNDSNS*

Description:

2-OXOISOVALERATE DEHYDROGENASE BETA SUBUNIT (EC 1.2.4.4)
 (BRANCHED- CHAIN ALPHA -KETO ACID DEHYDROGENASE COMPONENT BETA
 CHAIN (E1)) (BCKDH E1-BETA). - BACILL US SUBTILIS.

Assembly ID: 3858610

Assembly Length: 1078bp

> 3858610 Strep Assembly -- Assembly id#3858610
 CTAACCCTNGACGGGGCCGCTATCATCAGTCAAACAGCTAAAAATCTTGTCTGCAAAAGT
 CTCGATTAAGTGAAGCTTTTACAAAAGCCGTATTTCTGGAATAACTTGGAGATTGATCAT
 CTTATCCATCAATTCAGCCGATTCGATATTGTCTTTCAGCCAGTTGCAGACTTTTACGAT
 TGATTTTGGCAATTCGTAGACATAGGTGTTGTCTCTCAAAGGAATTTTGACAATACCTAA
 CTCCTTTGATATCTCGGGATACCGTCGCCTGAGTGGCAGTGATACCTGCTTCTTTCAAATG
 TTCTACAATTTCTTCTTGCGTGCCGATTTGATAATCTGTCCACCAATCTTCTAATTTTTC
 AAGTCTCTCTTTTTTATTCATTTTTTAAATTGACTATGCGCCCTCTCTACTGCTTCTTTAA
 TCTCAGCAAGAATCTGATTGCTTGCTGACTTTTCTTTTTTCAAATACACTAAAAATTCAA
 TATTTCCATGTCCACCTTGGATGGGAGAAAAGTCCAAGCCAAGGACTGAAAAACCTGCCT
 CTACTGCCATAGCTGTTACAGATTCAAGGACATTCTGATGAATCTTAGCATCTCGAATAA
 TTCCATTTTCCCAATCTGCTCACGTCCTGCCTCAAACCTGAGGTTTGACAAGTGCTACCA
 CCTGACCTTGATCAGCCAAGACACGGTGCAAGGCTGGCAAAATCAGACTAAGGGAAATGA
 AACTCACATCAATACTGGCAAAGCTCGGCTCCTGCTCGAAATCAGTCTTTTTCAGCATAGC
 GGAAATTGAACTGCTCCATGCTGACAACCTCGTGGGTCTTGGCGTAATTTCCAAGCCAAC
 GATTGGTACCAACATCGACTGCAAAGACCAACTTGGCACTATTCTGTAGCATGACATCGG

TAAAACCTCCAGTAGAGGCCCCGATATCAATCGTAGTCGCGCCATCCACCGACAAATCAA
AGACCTGCAAGGCCCTTTTCCAGTTTCAAACCACCACGGCTGACATACTTGAGTTTCTCC
CCCTTGAGTTTTAATTCGGTGTCTATCTGGAATTTCTCTCCTGGCTTGTCAAACCGTTC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 374 | 949 | R | 192 aa |

> 3858610-2 ORF translation from 374-949, direction R
VDGATTIDIGASTGGFTDVMLQNSAKLVFAVDVGTNQLAWKLRQDPRVVSMEQFNFRYAE
KTDFEQEPSFASIDVSFISLSLILPALHRVLADQGQVVALVKPQFEAGREQIGKNGIIRD
AKIHQNVLESVTAMAVEAGFSVLGLDFSPIQGGHGNIEFLVYLKKEKSASNQILAEIKEA
VERAHSQFKNE*

Description:

cytotoxin/hemolysin ORF2 tly - *Serpula hyodysenteriae*

Assembly ID: 3858716

Assembly Length: 928bp

> 3858716 Strep Assembly -- Assembly id#3858716

ACTTTCCTGACCTCTGTTTCCAAATAATCTTCCAAATGGACAGAGATCTACCGTTGTTTG
CATCGATAGCTGAGGTCTTTTTTTAGAAAATACCATCACTTTTAGAAAATATAAACACATT
TTTCGGATAAGATTAAGGTAAAAGCAGCTCGTTTATCCAGGGTCTGATGATGGTCTTCA
CGATAAACACATCCAATAACCAATGCATACTTTCTGCTGACCAATGACCTCGAACACTA
TGGCAAAAGGTCATCAACATCAAGCTTAAAGTTAAAGATAAAAATAGCGAACGTCTTGACT
TGTAATACCATCTCTATCAATAGTATTACGAGTCATTCCAATTCCACGCAATTATGCCA
TTTGGGATGGTTTTTGACACAACCACCTTAACATCAGAAGACACCCAGTATTCTCGAACTTC
AATCTATCCTCTTTCTATATTCTAACTGAAAGGACAATTCAATGATTCATTTAATAATGA
TTAGCGCCATTGCTCTAGCCATTGGAATTGGTTACCGCACCAAAATCAATATTGGCCTGC
TGGCTATTGCTTTTTCTTACCTCATCGCAACCACTCTCATGGGATTAAGTCCCAAAGAAC
TTCTTCATTTTTGGCCAACCTCACTCTTTTTTACCATTTTGTAGCGTCTCTCTCTTTTATA
ACGTTGCAACAATAACGGTACTCTTGATGTTTTGGCTCAACACATTCTCTACCGCACAC
GCACCCACCCTAACGCCCTCTACATGATTTTATACCTGATGGCAACCCTTTTGTCTGCTT
TAGGTGCTGGATTTTCACTACTATGGCCGTTTGCTGTCCTCTAGCGATTACCCTCTGTC
AAAAAGCGGACAAACACCCTTTGATTGGAGTCAAAGCGTCAATGGGAACCTTCAGGAAGGG
TAATTTGATAACCAAAGGAATAAAATTT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 238 | 402 | R | 55 aa |

> 3858716-1 ORF translation from 238-402, direction R
VSSDVKWLCQNHPKWHKLRGIGMTRNTIDRDGITSQDVRYFIFNFKLDVDDL*P*

Description:
unknown

Assembly ID: 3859124
Assembly Length: 847bp

> 3859124 Strep Assembly -- Assembly id#3859124
AAAAACGCACCATATCAAAAAGTTTGGATATCATGCGTCATGTCTTAACTAAT
TGACTATACTTTCTATTCAAATGAGCTTTTAACCAATTGATTGAGCCAATCCACTCTTAA
AACCAAAGGAGCAATTTCTCGGCTTAGCTGACTCTTCTCGGAATCTGAACCATGTACAAC
ATTTTGGATAATCTCATTTTCTCCAGCAGCTTTTGCAAAATCACCTCGAATAGTGCCTGG
TAAAGCTTCTTCTGGACGAGTTGCACCCATCATGGTCCGCCAAGTTTCGATTACTTTGGG
ACCAGAAATGACACCCACAAGAACTGGACCTGAAGTCATGAATTCACGAATCGGTGGGTA
AAAACCTCTGACCAACCAAGTCCTGATAGTGCTGGTCAATCAACTCTTCTGAAAACCTGTG
AACGAACTCCAATTTTTCGATTGTAAATCCACGTTGTTTCGATGCGCTTTAACTTCAC
CCACTAGCCCTCTTTTACACCATCTGGTTTGATGATAAAGAATGTTTGTTCATACCCG
TCTCCTTTGTCAGCTTCTTTCTTTTATTTTACCACATCTCGTGAAAAATGGAGAAAGTT
TTCAGAAGAGAGAATGAGAGAACCCTCGGGTTCTCTCATTCTCTCTTATTCTACTGTTTC
TTCCACAGTGTCACGGCAGTATCCACAACCTACTTCTGTTGTTTCTTCATTTCTTCTTC
CTCTACTGGAGGATTAAGGTATTCTTCTTCGTTGACAGCATGTGGTTCAAGGTTACGGTA
ACGGGCCATACCAGTACCAGCTGGGATGATCTTACCGATGAATAACATTTTCTTTAAAT
TCCAAGG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 73 | 453 | R | 127 aa |

> 3859124-1 ORF translation from 73-453, direction R
VDLQSKNWSFVHRFSEELIDQHYQDLVGQSFYPPPIREFMTSGPVLVGVISGPKVIETWRT
MMGATRPEEALPGTIRGDFAKAAGENEIIQNVVHGSDSEKSQLSREIAPLVLVRVDWLNQL
VKSSFE*

Description:

NUCLEOSIDE DIPHOSPHATE KINASE (EC 2.7.4.6) (NDK) (NDP KINASE)
(ABNORMAL WING DI SCS PROTEIN) (KILLER-OF-PRUNE PROTEIN). -
DROSOPHILA MELANOGASTER (FRUIT FLY) .

Assembly ID: 3859244

Assembly Length: 578bp

> 3859244 Strep Assembly -- Assembly id#3859244

ACAACCTAACTACCGNCTAATTCAGCGCGAACTTCTGCAGTAGCTGCTTCAACAACCTTCA
CGACGTGAAAGGATGAAGCGGTTTTCTTTAGCGTTAACTTCTTTGATTTTAGTATCAAAT
TCTTGACCTACAAAACGCTCAGCGTTACGTACGAAACGAGTATCCAACATTGAAGCTGGG
ATAAATCCACGAACACCTTCAAATTCTACTGAAAGTCCACCTTTAACGGCAGCGGTTCCCT
TTAACAGTAACAACCTTCTTCTTCGCGACCAACAAGTTTGTCCCATGCTTTGCGAGCTTCA
AGGCGTTTTTTTAGATGACAAGGTATGTAAGTGTATCAGTATCTTTACCAACTACTTGACG
AAGTACAAGAACATCCAATACTTCTCCTACTTTAACAAAAGTCATTGATATCTGCATCACG
ATCGTTTGTCAATTCGCGAAGAGTCAAGACACCCTTCAACACCAGTTCCCAGAAGAATGC
AACGTTAGCTTGAGTCGCATCAACTGTCAATACTTCAGCACTAACACATCACCAGTCTCA
ACTTGACTNACGCTATTGAGCANATCTTCAAATTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 310 | 462 | R | 51 aa |

> 3859244-2 ORF translation from 310-462, direction R
VLKGVLTRELTLNDRDADINDFVKVGEVLDVLRQVVGKDTDTVTYLVLI*

Description:

unknown

Assembly ID: 3859250

Assembly Length: 888bp

> 3859250 Strep Assembly -- Assembly id#3859250

GTAGTTATAGTAGGGGTCGGATTGAAATGCCACNGCGCTTCTTGGAGTTTCTGATACCGT
TTAAAATAGCGTTGGGCATTCTGGTTGGGAGTCAGAGCCTTATCAAGCGCAATCATGATA
GGTTGGTTGGTATAGTAGTTGTCTAGGATAACCTGGTTCTTGGTCGTTAGGCACCTGGTG
GAGGAAGGTTGTCAGCAATTCTCCTTTTTGACGAAATCTTCAGCGTTGTCTGTGCGCCAG
TAACTATTTTTCTGTTTTTTGAGTTTGTGTGCGGTTTTCTGAAGTTCATTTTCAACACG
ACGAATCAGTTCACTGGCCTGCTGTTTGACGCGGTGCGGCTCAGCCTTATCCTTATAGTA
GGTGTCCAACAAATCAGAAAGATTTGCAAAGGCTCTCCACCTGATTTGCAAAGGAAC
TGGACTGAAGGAAGTCTCAGTCAAGCATGGCTTGGTTTCTGATTGAAAAAATTTTCGGAA
AGCGGAAAGTTTTTCACTAACCAGTATCCTTTCCAATTCATTTGCCGTATCGCGTCCCAG
ACCTTGAAAGAGGCTTTGAAGATTTTTGCTGTTAGTTCTTGGGTTTGCAGGATTTCAA
GAGCTTTTCATCCTTGATAGTAAAAGGATTGAGAGATTCTGTACTTGGCGGAGCGATATA
GGTCGATCCTGGAAGTAAGGTGCGGTAGCTATTTTGTGAAAAGCCGACGTGTTTGATAAC
TTTCGAGGATTTTATGACTGCTTTTATCCGACCAGTTAGAATATTACTGTGTTTCCCCATA
ATTTGATAATCAAGGTAGCCTGGATATGGTCTCCAATCTCGTTTTTATTGGAACTGTA

ATTTCCACAATACGGTCATTTTCCACTTGCTCAATCGACTCAATCAGG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 244 | 402 | R | 53 aa |

> 3859250-1 ORF translation from 244-402, direction R
VGEPFANLSDLLDTYYKDKAERDRVKQQASELIRRVENELQKNRHKLKKQEK*

Description:

STRFBP5A NCBI gi: 496253 - Streptococcus pyogenes.
Fibrinogen/Fibronectin binding protein

Assembly ID: 3859588
Assembly Length: 513bp

> 3859588 Strep Assembly -- Assembly id#3859588
ATCGAATTTTGTCTTTCATAGAGAGCTACCTGAGTTCTATTCAAGCTCAGGTAGTACTT
TCTTATAAACTAGACAACTAACTGTCATTCTACCATCAGATTACAAGACATCATCGTCA
CTCACCTTGGAATTCAATGTCGTACCCCAATGGGTAATTTTACGGTGGGGTTGAGCTAAA
ATTGGTCTGTTTTTCATAGATTGTTTGCCATCTATTCCATAGTAGGCCCGTCTTTTTCTCA
ATCTTAACTCGCAGATTTCTCATATTTTCTTTGATTGGGAGGTTGAGGACAAAACCTGCA
GTCTGGTTGCGACCGTTTCCTTCCCAAGAATGACTACGAACAACCTTGGTTTCCATCTTTA
TCTACTGGAACCTTCTTCCCAAGTTATGGAGTAGCGGGCAATGTAAGCTCCACTGTGTTGA
ATTATCAATGTTTTATCTTTACAGGGAGTCTGACTGATTGGTTGAACTGGCTTAGAAAC
TTGTGTCGCCGTTTCAGCATTCGTAGCTATAAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 102 | 443 | R | 114 aa |

> 3859588-1 ORF translation from 102-443, direction R
VKDKTLIIQHSGAYIARYSITWEEVPVDKDGNOVVRSHSWEGNGRNQTAGFVLNLPIKEN
MRNLRVKIEKKTGLLWNRWQTIYENRPILAQPHRKITHWGTTLNSKVSDDDVL*

Description:

PNEUMOLYSIN (THIOL-ACTIVATED CYTOLYSIN). - STREPTOCOCCUS
PNEUMONIAE.

Assembly ID: 3859774
Assembly Length: 214bp

> 3859774 Strep Assembly -- Assembly id#3859774
 ATCGAATTCTAACATGTGCTTCTCCTTCTATTGTTCCCTATCTTTAAAATCTACTCCTTCA
 TGCTCCAAGAGCCAAGCTTTCTTTTCCACTCCTGCAGCATAACCTGTCAGACGCTTGCCCT
 GCTCCCAACACACGATGACAAGGTACTAGGATAGACCAAGGATTGCGTCCCCTGCTCCA
 CCAATTGCTTGAGCAGAAGCCACTTGCAGGTCTT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 9 | 131 | R | 41 aa |

> 3859774-1 ORF translation from 9-131, direction R
 VLGAGKRLTGYAAGVEKKAWLLEHEGVDFKDRNNRRRSTC*

Description:

GLUTAMATE RACEMASE (EC 5.1.1.3). - ESCHERICHIA COLI.

Assembly ID: 3860140

Assembly Length: 1084bp

> 3860140 Strep Assembly -- Assembly id#3860140
 CTCCAGCAATGGATCCAAGTATGATGGGCGGGATGATGTAAGCTTTCTATAGAAAACACC
 TTATAAAAAACACGAAAGGAGGGAATGACTAACCTTCTTTTTATAATATTTACTTCTAA
 GATTGATGGTGAGCTCTCCTAACTTATATGATAAAATAAGACTAGAGGAAAGGAGAAGAA
 CATGATCGATGTACAAGAAATTCTGTGCAAGATGACCCCAATCAGAAGATTAATTATGA
 CCGTGTCATGCAGAAAATGGTACAAGCATGGGAAAAAATGAGTAGCGGCCAACCATTCT
 CGTGCATGTTTGCTGTGCCCCCTTGTAGTACCTATACACTAGAATATTTGACCAAGTATGC
 AGATGTGACCATCTATTTTGCCAATTCTAATATCCATCCCAAGGCAGAATACCATAAGCG
 GGTCTATGTCACCAAGAAATTTGTTAGTGATTTTAATGAGCAGACAGGAAATACGGTTCA
 GTACCTAGAAGCTCCCTACGAACCCAATTAATACCGAAAACCTAGTTAGGGGGCTAGAGGA
 GGAGCCCGAAGGTGGCGACCGTTGCAAGGTTTGTGTTTACTACCGACTGGATAAAACAGC
 GCAAGTGGCTATGGACTTGGGCTTTGACTACTTTGGTTCAGCCTTGACCATCAGTCCCTCA
 TAAGAATTCTCAAACCTATCAATAGCATCGGAATCGATGTGCAAAAAATTTACACGCCCCA
 CTATCTTCCCAACGATTTCAAGAAAAATCAAGGCTACAAACGTTTCAGTAGAGATGCGTGA
 GGAGTATGATATCTATCGTCAATGTTATTTGTGGCTGCGTCTATGCAGCCCAAGCCCAGAA
 TATTGACCTGGTTTAAGTTGAGTAGGACGCCACAGCATGCTTGCTGGATAAGGATGTTGA
 GAAAGACTATTCTCATATCACATTTATAGTAGATTGAACTAGAAATAGTACACCTTTACT
 TCTCAAACATTGTTAGAAATCGATTGCGCTGTCTTATTTTCAATTTTAATACTGGTACG
 AAATTAGATATATCAATGATAACTTGCCTCAAGGTAGGTTTTTTGATAGTAGAAAAGCGA
 TAGA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 302 | 511 | F | 70 aa |
| 2 | 605 | 856 | F | 84 aa |

> 3860140-1 ORF translation from 302-511, direction F
 VHVCCAPCSTYLTLEYLTKYADVITYFANSNIHPKAEYHKRVYVTKKFVSDFNEQTGNTVQ
 YLEAPYEPN*

Description:
 unknown

> 3860140-2 ORF translation from 605-856, direction F
 VAMDLGFDYFGSALTISPHKNSQTINSIGIDVQKIYTPHYLPNDFKKNQGYKRSVEMREE
 YDIYRQCYCGCVYAAQAQNIDL*

Description:
 unknown

Assembly ID: 3860206
 Assembly Length: 1124bp

> 3860206 Strep Assembly -- Assembly id#3860206
 ATCGAATTTCATTGACTGCCTGAAAAGACTTCAACTCGTCTGCCTGATAACCGAAAGACTT
 GGTTACTTTGATACCTGATACGGACTCCTGTACCTTGTTATTGAGTTCAGAAAAAGCAGC
 TTGGGATTTCGCCAAAGGCCTTATGAGTCTTTCTCCCTAGGCGACTAGTCGTATAGGCCAT
 GAAAGGTAGGGGGAGAATGGCAACAAGAGTCATCTGCCATGAGATGCTAAAGAGCATGGT
 CAACAAAGTCACCAGAGCCGTGATAGAGGCATCCACCGCAGACATGACACCGCCACCTGC
 TAAACGAGTCAAGGAATTGATATCATTTGGTTGCGTGTGCCATCAGATCACCCGTCCGATA
 GGTTTGATAAAAGGCTGACGACATTTTTGTGAAATGCTTAAACAAGCGAGACCGCATGAT
 CTGTCCCAAGCAATAAGAGGTCCCAAGGATATACATACGCCACACATAGCGCAAATAGTA
 CATACCAAAGGCTGCAAGTAGCAAGTAAATAGGCTAAGAAGGAGGTCCTGCTGGGTAA
 TTGCCCCGATGTGATGGCATCAATAACCCGCCCCATAACCATAGGAGGAATGAGATTGAG
 GACGGAAACCAAGACCAGGGCCACAATCCCGACTAGATAACGGCGTTTTTCTAACTTGAA
 AAACCACCAAATTTTTGAATAATGGACATAAAATCCCTTTCTGGATTGCAAATAGAAAC
 CTGAGGCCAATACTCAATGGAAAATCAAAGAGCAAAGTGGAACTAGCCGCAGGCTGCT
 CAAAGCACTGCTTTGAGGTTGTAGATAGAAGTACGAAGTCAGTAACCTACATACGGCAA
 GGCGACGTTGACGCCGTTTGAAGAAATTTCCGAAGAATACAAGACCCCAGGTTTTTCTTA
 TTTATAAGTTACCACTGTAACAGCACCTTGTCATATTCAGCAATAAAGATATTGGCTAC
 ATTGTCATGCCCTTGTTTACTGAGGTTATCAAGCAACCACTCCTCGCTACGAACAATCGA
 TCCCAAGACATCTACTTGAATCACACCGTCAGTCACAAGTGGATACTTAGGATTTTCATC
 TCCCATTTGCACAACGATGAGTTGCCCATTTTGCTCTTGACACAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 898 | 1056 | R | 53 aa |

> 3860206-2 ORF translation from 898-1056, direction R
 VTDGVIQVDVLGSIVRSEEWLLDNLKQGHNDNVANIFIAEYDKGAVTVVITYK*

Description:

unknown

Assembly ID: 3860270

Assembly Length: 1242bp

> 3860270 Strep Assembly -- Assembly id#3860270

TTACCTTCATTGCAGCCATTATTGGTTCTTGTGTGCAGCCAGATTTTAAGTATTCTTTATA
 AGACACCTGCTGTGGTCTTTATCTTGGCCATTTTGGCACCGCTGGTTCCAGGTATCTCT
 CCTACCGAACAACCTGCCTTTTTTGTGACAGGGGACTATAATAAAGCACTGGCAAGTGCGA
 CCTTGGTTGTCATGTTGGCTTTGGTAATCTCTATTGGAATGGCTAGCGGAACAGTGATTC
 TCAGACTGTATCATTTATATAAAAACACATCGAGTATCGTAGACTTTACAGAAATAAAAGA
 ATTTTCTGAAAAATGAGATAAATAAATTAACAACGCTTCTATATGTGCGAGAATACCGC
 ACTTATGAAGAAATTGCGGCTGATTTTGGTATCCACGAAAGCAACTTAATCCGTCGGAGC
 CAATGGGTTGAAGTAACTCTTGTTCAAAGTGGTGTACGATTTCAAAAACATCTTAGT
 GCTGAGAATACGGTGATTGTGGATGCAACAGAGGTAAAAATCAATCGCCCTAAAAACAA
 TTAGCGAATGATTCTGGTAAAAAGAAATTTACGCTATGAAGGCTCAGGCGATTGTCACA
 AGTCAAGGGAGAATTGTTTCTTTGGATATCGCTGTGAACTATTGTCATGATATGAAGTTG
 TTCAAATGAGTCGCAGAAATATCGGACAAGCTGGAAAAATCTTGGCTGATAGTGGTTAT
 CAAGGGCCCATGAAGATATATCCTCAAGCACAACTCCACGTAAATCCAGCAAACCTCAAG
 CCGCTAATAGCTGAAGATAAAGCTTATAACCATGCGCTATCCAAGGAGAGAAGCAAGGTT
 GAGAACATCTTTGCCAAAGTAAAAACGTTTAAATGTTTTCAACAACCTATCGAAATCAT
 CGTAAACGCTTCGGATTACGAATGAATTTGATTGCTGGCATTATCAATTATGAACTAGGA
 TTCTAGTTTTTGCAGGAAGTCTATTATTTTCCTTATTGTCTGTAAGTCTACTGACCTTGTT
 GTTTATCCCAGTCATGGTTTCTAGTTCGGGCTCAGAGTTTCAAAGTGGATGGCAAGAGCA
 TCAATTGATTGCTGAGAAGGTTAGTAAAAACCTTGACAAGACATTTGATAAGGATGTCAG
 AAAAATTCCGACCAGTCAGTTTTATCAAAAATTTGTAGATGAGATGGGAAGGATTTACTC
 AGGAAATTTGATCCTCCCAGGAGCTGATAACTGTGAATGGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 346 | 966 | F | 207 aa |

> 3860270-1 ORF translation from 346-966, direction F

VREYRTYEEIAADFGIHESNLIRRSQWVEVTLVQSGVTISKTHLSAENTVIVDATEVKIN
RPPKQLANDSGKKKFHAMKAQAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQAGKIL
ADSGYQGPMKIYPQAQTPRKSSKLKPLIAEDKAYNHLSKERSKVENIFAKVKTFKMFST
TYRNRKRFRGLRMNLIAGIINYELGF*

Description:

ISL2 protein - *Lactobacillus helveticus* (Probable transposase)

Assembly ID: 3860438

Assembly Length: 1575bp

> 3860438 Strep Assembly -- Assembly id#3860438

GTGATGGGGCCTCAGGGAAATGGTTTTGACTTGTCTGACCTTGATGAGCAGAATCAGGTT
CTCCTTGTGGTGGTGGGATTGGTGTTCACCCTTGCTTGAGGTGGCCAAGGAATTGCAT
GAACGTGGAGTGAAAGTAGTGACAGTCCTCGGTTTGTCTAATAAGGATGCTGTTATTTTG
AAAACGGAATTGGCTCAGTATGGTCAGGTCTTTGTAAACGACAGATGATGGTCTTATGGC
ATCAAGGGAAATGTTCCGTTGTTATCAATGATTTAGATAGTCAGTTTGATGCTGTTTACT
CGTGTGGGGCTCCAGGAATGATGAAGTATATCAATCAAACCTTTGATGATCACCCAAGAG
CCTATTTATCTCTGGAATCTCGTATGGCTTGTGGGATGGGAGCTTGCTATGCCTGTGTTC
TAAAGTACCAGAAAGCGAGACGGTCAGCCAACGCGTCTGTGAAGATGGTCCTGTTTTCC
GCACAGGAACAGTTGTATTATAAGGAGAAAATTATGACTACAAATCGATTACAAGTGTCT
CTACCTGGTTTGGATTTGAAAAATCCGATTATTCCAGCATCAGGCTGTTTTGGCTTTGGA
CAAGAGTATGCCAAGTACTATGATTAGACCTTTTAGGTTCTATTATGATCAAGGCGACA
ACCCTTGAACCACGTTTGGGAATCCAACCTCAAGAGTGGCAGAGACGCCCTGCTGGTATG
CTCAATGCAATTGGCTTGCAAAATCCTGGTTTAGAGGTTGTTTTGGCTGAAAAGCTACCT
TGGCTGGAAAGAGAATATCCAAATCTTCCTATTATTGCCAATGTAGCTGGTTTTTCAAAA
CAAGAGTATGCAGCTGTTTCTCATGGGATTTCCAAGGCAACTAATATAAAAGCTATCGAG
CTCAATATTTCTTGTCCCAATGTTGACCACTGTAATCATGGACTTTTGATGGTCAAGAT
CCAGATTTGGCTTATGATGTGGTGAAAGCAGCTGTGGAAGCCTCAGAAGTGCCAGTTTAT
GTCAAATTAACCCCGAGTGTGACCGATATCGTTACTGTGCGAAAAGCTGCAGAAGATGCG
GGAGCAAGTGGCTTGACTATGATCATACTCTGGTGGGATGCGCTTTGACCTCAAAACCAG
AAAACCAATCTTGGCCAATGGAACAGGTGGAATGTCAGGTCCAGCAGTTTTCCAGTAGCC
CTCAAACCTCATCCGCCAAGTAGCCCCAAACAACAGACCTGCCATCATTTGGAATGGGGGGA
GTGGATTCGGCTGAAGCTGCCCTAGAAATGTATCTGGCTGGGGCATCTGCTATCGGAGTT
GGAACAGCTAACTTTACCAATCCTTATGCCTGCCCTGACATCATCGAAAATTTACCAAAA
GTCATGGATAAATACGGTATTAGCAGTCTGGAAGAACTCCGTCAGGAAGTAAAAGAGTCT
CTGAGGTAAACTGCAATCAATCTGTTCTTGATTTTTTATTAGTTTGTAATATGAATTTAG
GAGAATTTTGGTACAATAAAATAAATAAGAACAGAGGAAGAAGGTTAATGAAGAAAGTAA
GATTTATTTTTTTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-------|-----------|--------|
| ----- | ----- | ----- | ----- | ----- |

| | | | | |
|---|-----|------|---|--------|
| 1 | 1 | 276 | F | 92 aa |
| 2 | 460 | 1128 | F | 223 aa |

> 3860438-1 ORF translation from 1-276, direction F
 VMGPQGNQFDLSDLDEQNQVLLVGGGIGVPPLLEVAKELHERGVKVVTVLGFANKDAVIL
 KTELAQYQGVFVTTDDGSYGIKGNVPLLSMI*

Description:
 unknown

> 3860438-3 ORF translation from 460-1128, direction F
 VKMVLFSAQEQLYYKEKIMTTNRLQVSLPGLDLKNPIIPASGCFGFGQEYAKYYDLDLLG
 SIMIKATTLEPRFGNPTPRVAETPAGMLNAIGLQNPGLVLAELPWLEREYPNLPPIA
 NVAGFSKQEYAAVSHGISKATNIKAIELNISCPNVDHCNHGLLIGQDPDLAYDVVKA
 ASEVPVYVKLTSPVTDIVTVAKAAEDAGASGLTMIILWWDAL*

Description:
 DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE
 OXIDASE) (DHODEHASE). - BACILLUS SUBTILIS.

Assembly ID: 3860544
 Assembly Length: 776bp

> 3860544 Strep Assembly -- Assembly id#3860544
 CTAAGATATCAGAATAACAACGAAATCGAAGCATTAACAAATATTACTTCTAAGAAT
 AGCGAGATTGATAGTCAACAAAGCAATATTAAGGATATGACCGTACCTATAATGATCCAA
 CTTCTCAGGCTTATAATATTTATGCTCAATTAATTAGTGAGTTAGGTACTGCTCGTTCAA
 ACAACAATAAAAGTATTACAGAGCTTGAGGCTAATCTTGAGGTGGCAACAGGTCAAGATA
 AAGCTCATAGTATATTAGCGTCAAATGAAGGTACTCTGCATTATCTGGTACCTTTGAAAC
 AAGGAATGTCTATTCAGCAGGGGCAAACGATAGCAGAAGTTTCAGGGAAAGAAAAAGGTT
 ACTATGTAGAGGCTTTTGTACTTGCGAGTGATATTTCTCGTGTTTCAAAGGAGCAAAAG
 TTGATGTTGCTATTACTGGTGTGAATAGTCAAAAATATGGAACACTAAAGGGACAAGTCA
 GACAGATTGATTCAGGAACAATTTCCCAAGAAACGAAAGAGGGGAATATTAGCCTCTATA
 AAGTCATGATAGAATTAGAAACCTTAACCTCTAAAACATGGAAGCGAGACGGTCATACTCC
 AAAAGGATATGCCAGTTGAAGTGCGGATTGTCTATGATAAAGAAACCTATCTTGATTGGA
 TTTTAGAAAATGTTAAGTTTCAAGCAATAATTGGTTTTTAAACCTTAGGTAACCTATAAAAA
 CAAATAAGGTAGAGAAAGGATATTTTATCTAAGTTAGCTCACATTACTGCCATTCC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 222 | 689 | F | 156 aa |

> 3860544-1 ORF translation from 222-689, direction F
 VATGQDKAHSILASNEGTLHYLVPLKQGMSIQQGQTIAEVSGKEKGYVVEAFVLASDISR
 VSKGAKVDVAITGVNSQKYGTLLKGQVRQIDSGTISQETKEGNISLYKVMIELETLLTKHG
 SETVILQKDMPEVRIVYDKETYLDWILEMLSFKQ*

Description:

unknown

Assembly ID: 3860558

Assembly Length: 1487bp

> 3860558 Strep Assembly -- Assembly id#3860558

CTGGCCTTTCTCCACCAAATTGTTCCCTTGAGGGAAGGAAGTCAGAACACTAGCCGTTGC
 ATCTTCCTTTTGCTTTTCAATCGTAATTCCAGATAATTTTCCCATTCTTTTGGTGACC
 CCGGGAGGCAGGATTGAATGGCTTGAGGGAATGACAACTTGTCTAGCAAGAATGGTC
 AAGGCACCTCCGTCTACAATCAAAATCTGATTGTTGGGCTTAAATTAACAAAGACCTGTTTT
 ACTAGATTTTTCTCCAGAAGCATCGTCTCGTAAACCAGGCCCCAGCAAGATAACTTCTGCC
 TTCTCCAATTGCTCTTTTAACAATTGCTGGTCTTGAAGAGAAAAGGCCATAGGCTCAGGT
 AAATGGCTGTGCAGAGCCGGGATATTTTCCCTGTCCGTCCAACGGTCACCAATCCTGCA
 CCGCTTTTTTACAGCTGCTAAAGCAGCCATGATGATGGCACCTCCATAAGGATAAGTACCA
 CCAAGCAGCAGCAGACGACCATAATCTCCTTTATGACTTGAACGAGAACGTTCAATAATA
 ACTTTTTCTAGTAAGGTTTGATTAATCACTTTCATCCTTTTTCCCTCTCACTTTTATTAT
 ACAACAAAAGGAGACGCAGACCTCCTTTTGTAATCTTATATCTAAAATTTAATATTCAT
 TTCTGCCATTTTAGATATAGCTATAGAAAATACACTCTATTAATCGAATGTTTCTCTTAT
 TTTCTATCCAATGTCCGAAGTGCTGCTTGATAAGTTTGCTCCATCAGCATGGTAATGGTC
 ATAGGACCGACACCTCCAGGGACTGGCGTGATATGGCTAGCAAGTGGTGCAACTGCCTCA
 TAATCAACATCTCCACAGAGCTTCCCATTTTCATCTCGGTTTCATCCCAACGTCAATGACA
 ACCGCACCTGGTTTGACAAAGTCAGCAGTCACAACTTGGCGCGGCCGATTGCGACTACA
 AGAATATCTGCTTTAGCAGCCACCTTGGCAAGATTATGAGTTTCGTGAGTGGGCCAAGGTT
 ACTGTGCGATTTTGTAGCCAAAAGAAGCTGAGCCATAGGTTTTCCAACGATATTTGAACGA
 CCGATTACGACCGCATTTTACCTTCCAAGTCAATCCCATATTCATGAAACATTTCCATA
 ATTCCTGCAGGTGTGAGGGAATCATGACTGGATGTCCAGACCAAAGACGTCCCATGTTT
 AGGGGATGGAAACCATCCACATCCTTTTCTGGGTCAATGGCTAATAAAACCGCCTCTTCA
 TCGATATGTTTTGGTAATGGCAACTGGACCAAAATCCCATGCCAAGCTGGATCCTGATTA
 TATTTAGCAATCAGGTCTAACAATTCCTCTTGAGTAATGGTCTCTGGAACCTGCACTACT
 TCGGTACGGGAACCGCCGCAAGAGCTGACCTCTCCTTGTTGCGAACGTTAAACTTGGCT
 GGCTGGATTATCCCCAACCAAAATCACTACCAAAACCAGGCACTAGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 717 | 1376 | R | 220 aa |

> 3860558-2 ORF translation from 717-1376, direction R
 VRVPETITQEELLDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKDVDGFHPLN
 MGR LWSGHPVMIPSTPAGIMEMFHEYGIDLEGKNAV VIGRSNIVGKPM AQLLLAKNATVT
 LAHSRTHNLAKVAAKADILVVAIGRAKFVTADFVKPGAVVIDVGMNRDENGKLCGDVDYE
 AVAPLASHITPVPGGVGPM TITMLMEQTYQAALRTLDRK*

Description:

5,10-methylene-tetrahydrofolate dehydrogenase (fold) homolog -
 Haemophilus infl uenzae (strain Rd KW20)

Assembly ID: 3860568

Assembly Length: 1634bp

> 3860568 Strep Assembly -- Assembly id#3860568

CGTGCCTTGGCCAATGATCCAAAAATCTTGATTTTCAGACGAGTCGCTTCAAATTTTCGGCC
 CCTGGACCCTTAAGACCAACCCAAGCAGATTTTGGCCCTTGGTTGCAAGATTTGAACCAA
 AAATTAGGCTTGA CTGTTGTCCTGATTACGCATGAAATGCAGATTGTCAAAGACATTGCC
 AACCGTGTTCAGTTATGCAGGATGGGCATTTGATTGAAGAGAGTAGTGTGCTTGAAATC
 TTCTCAGACCCTAAACAACCTTTGACTCAAGACTTTATCTCAACAGCTACAGGTATTGAC
 GAAGCCATGGTCAAATCGAGAAGCAAGAAATCGTGGAACTTGTCTGAAAACAGTCTC
 TTGGTGCAACTCAAGTACGCTGGATCTTCAACAGACGAGCCACTTTTGAATGAATTGTAC
 AAGCATTATCAAGTAATGGCTAATATTCTCTATGGGAATATCGAAATCCTCGATGGTACT
 CCTGTTGGAGAATTGGTGGTGGTCTTGTTCAGGTGAAAAAGCAGCGCTGGCAGGTGCTCAA
 GAAGCCATTCGTCAAGCAGGCGTACAGTTAAAAGTATTGAAGGGAGGACAGTAAGATGGA
 ATCATTGATTCAAACCTATTTACCAAATGTCTATAAGATGGGTTGGTCTGGTCAGGCAGG
 CTGGGGAACAGCTATCTACCTAACCTCTATATGACAGTTCTTTTCCTTCATTATCGGAGG
 CTTCTTGGGGCTAGTGGCAGGTCTCTTTCTCGTCTTGACAGCGCCAGGTGGTGTCTTGGA
 GAATAAAGTCGTATTCTGGATTTTAGACAAAATTACCTCAATTTTTCGTGCGGTTCCCTT
 TATCATCCTCTTGGCAATCTTGTCAACACTTTCTCACTTGATTGAAAAACAAGTATCGG
 GCCAAATGCAAGCCCTTGTCCCACTTTCTTTTGCAGTCTTTGCCTTCTTTGCCCGTCAGG
 TGCAGGTTGTCTTGGCTGAAATGGATGGCGGTGTCATTGAGGCGGGCTCAAAGCGAGCGG
 AGCGACTTCTGGGACATCGTGGGTGTTTACCTATCAGAAGGTCTTCCAGATTTGATCCG
 TGTGACGACTGTGACCTTGATTTCCTTGTTGGGGAAACAGCTATGGCCGGTGCGGTG
 AGCTGGTGGTATCGGTAACGTAGCCATCGCTTATGGATTTAACCGCTACAATCACGATGT
 GACCATCTTGGCAACCATCGTTATCATTTTGATTATCTTTGCAATCCAATTCTTAGGAGA
 TTTCTTGACTAAGAAATTGAGCCATAAATAAAAAAGAGCCGTGTGGCTCTTTTAACTGA
 TCAGATTTTCTGGGCAAATTTTTTACTCAAGGCTTGTCCAATCAAGGCACCCACTAGGGC
 TCCGATGACAATACTTGCGATAAATAGAAGGACAGTTCAGGGTTTGGAGCGACCATGAT
 GCGGTGATATATTCTTGGGATTTTCCCTCTTGCCAGAAGAGTAGCCATATAGGCTTTGGG
 CGCAATCCACATAAGCAAGATTGGTCTGTGTACTAAAGGCGAAAATAATGAAAGAAAG
 GAAGTTCTTTGTTTTGTCCTTGTATTTTCTTAAATGAGCTACTCCATCTGCTAGGAGGCC
 ACAGATAATTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1040 | 1291 | F | 84 aa |

> 3860568-3 ORF translation from 1040-1291, direction F
 VGVYLSEGLPDLIRVTTVTTLISLVGETAMAGAVGAGGIGNVAIAYGFNRYNHDVTILATI
 VIILIIFAIQFLGDFLTKKLSHK*

Description:

unknown

Assembly ID: 3860582

Assembly Length: 1087bp

> 3860582 Strep Assembly -- Assembly id#3860582

GGAATCATGATGATGTCACTGCTAAATGGTTTCTTAGAAAAATATTTCTGAGCGCTTA
 CAGATTAGTTTGGGCTTGCTGATTTTATCATTGAGCGGTACAGCTCCCTTCTGGTACCAA
 GCCTATCCCTTTGTCTTTGGAACACGGCTTCTCTTTGGTTTGGGTCTTGGGATGATCAAT
 GCCAAGGCCATTTCTATTATCAGTGAACGCTACCAAGGAAAAAGGCGAATTCAGATGTTA
 GGGCTACGCGCTTCTGCAGAGGTCGTTGGAGCTTCTCTCATTACCTTGGCCGTCGGTCAA
 GTTGTTGGCCTTTGGTTGGACAGCTATCTTTCTAGCCTATAGTGCTGGATTTTGGTGCT
 GCCCCTTTATCTGCTCTTTGTCCCTTATGGAAAATCAAAGAAAGAAGTCAAGAAAAGAGC
 GAAGGAAGCAAGTCGTTTAACTCGAGAAATGAAAGGCTTGATTTTTACCTTAGCTATCGA
 AGCGGCAGTTGTAGTTTGTACCAATACAGCTATTACCATCCGTATTCCAAGTTTGATGGT
 GGAAAGAGGATTGGGGGATGCCCAGTTATCTAGTTTTGTTCTTAGTATCATGCAGTTGAT
 CGGGATTGTGGCTGGGGTGAGTTTTTCTTTCTTGATTTCTATCTTTAAAGAGAACTGCT
 CCTCTGGTCTGGTATTACCTTTGGCTTGGGGCAAATCGTGATTGCCTTGCTTCATCCTT
 GTGGGTGGTAGTAGCAGGAAGTGTCTGGCTGGATTTGCCTATAGTGTAGTCTTGACGAC
 GGTCTTTCAACTTGTCTCTGAACGAATTCCAGCTAAACTCCTCAATCAAGCAACTTCATT
 TGCTGTATTAGGCTGTAGTTTCGGAGCCTTTACGACCCCATTCGTTCTAGGTGCAATTGG
 CTTACTAACTCACAATGGGATGTTGGTCTTTAGTATCTTAGGAGGTTGGTTGATTGTAAT
 CTCTATCTTTGTTCATGTACCTACTTCAGAAGAGAGCTCTAGGATTGATTCCTAAGTTTTT
 CTTTGTGATACTCAATGAAAATCAAAGAGCAAACATAGTTGATTGAGTTTGAATAGTAT
 GCTGTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 356 | 1027 | F | 224 aa |

> 3860582-1 ORF translation from 356-1027, direction F
 VLPLYLLFVPYGKSKKEVKKRAKEASRLTREMKGFLIFTLAIEAAVVVCTNTAITIRIPSL

MVERGLGDAQLSSFVLSIMQLIGIVAGVSFSFLISIFKEKLLLWSGITFGLGQIVIALSS
 SLWVVVAGSVLAGFAYSVVLTTFVQLVSERIPAKLLNQATSFVVLGCSFGAFTTPFVLGA
 IGLLTHNGMLVFSILGGWLIVISIFVMYLLQKRALGLIPKFFF*

Description:

unknown

Assembly ID: 3860724

Assembly Length: 1191bp

> 3860724 Strep Assembly -- Assembly id#3860724

GGATTCCAACGATTATGAACTTGACTGGTCCACTGATTCATCCAATGGCTTTAGAAACAC
 AGCTTTCTTGGAAATTAGTCGTCCAGACTCCTAGAAAGTACAGCTCAGGTTTTGAAAATAT
 GGTGCGAAACGTGCCATCGTGGTTGCTGGACCAGAAGGGTTGGATGAAGCTGGCTTGAAC
 GGAACAACCNAGATTGCACTTNTTGAATAATGGCGAAATCAGCTTGTCAGCTTTACTCCA
 GAGGATTTGGGAATGGAAGGCTATGCTATGGAAGATATTCGTGGTGGGAATGCTCAGGAA
 AATGCAGAAATTTTGCTTAGCGTTCTGAAAAACGAAGCAAGTCCATTCTTGGAACGACA
 GTCTTGAATGCTGGTCTTGGTTTCTATGCTAATGGTAAGATTGATAGCATCAAGGAAGGA
 GTTGCCTTGGCCCGTCAAGTGATTGCTAGAGGCAAGGCCCTTGAAAACTCAGACTGTTA
 CAGGAGTACCAAAAATGAGTCAGGAATTTTGTAGCACGAATCTTAGAGCAGAAGGCGCGTG
 AGGTGGAGCAGATGAAGCTGGAGCAAATCCAGCCTCTGCGCCAGACCTATCGCTTGGCAG
 AATTTTTGAAGAATCATCAGGACCGCTTGCAGGTAATCGCTGAGTCAAGAAAGCTAGCCC
 TAGTTTGGGAGATATCAATCTCGATGTGGATATTGTGCAACAGGCCCAGACTTATGAAGA
 AAACGGAGCAGTGATGATTTTCGGTGTGACAGATGAGGTTTCTTTAAAGGGCATTGGA
 TTATCTACGGGAAATTTCCAGTCAGGTAGAGATTCCGACGCTCAACAAAGACTTTATCAT
 AGATGAAAAGCAAATCATCCGCGCTCGCAATGCAGGTGCGACAGTTATCTTGCTTATGT
 GGCAGCCTTGTCCGAAGAACGCCTCAAGGAAGTGTATGACTACGCGACAGAGCTTGGTCT
 GGAAGTCTTAGTGGAGACTCACAATCTAGCTGAACTAGAGGTAGCCACAGACTTGGTGG
 CTGAGATTATCGGGGTCAACAACCGCAACTTGACTACCTTTGAAGTCGACTTGCAGACCA
 GTGTAGATTTAGCCCCTTACTTTGAGGAAGGTCGCTATTACATTTCTGAATCTGCCATTT
 TCACAGGGCAGGATGCGGAACGACTAGCCCCATACTTTAACGGAATTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 139 | 498 | F | 120 aa |
| 2 | 686 | 1024 | F | 113 aa |

> 3860724-1 ORF translation from 139-498, direction F

VVAGPEGLDEAGLNGTTXIALXENGEISLSSFTPEDLGMEGYAMEDIRGGNAQENAEILL
 SVLKNEASPFLETTVLNAGLGFYANGKIDSIKEGVALARQVIARGKALEKLRLQLQEQK*

Description:

ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18). -
 LACTOCOCCUS LACTIS (SUB SP. LACTIS) (STREPTOCOCCUS LACTIS).

> 3860724-2 ORF translation from 686-1024, direction F
 VDIVQQAQTYEENGAVMISVLTDEVFFKGHLDYLREISSQVEIPTLNKDFIIDEKQIIRA
 RNAGATVILLIVAALSEERLKELYDYATELGLEVLVETHNLAELEVAHRLGG*

Description:

INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE (EC 4.1.1.48) (IGPS). -
 LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3860858

Assembly Length: 858bp

> 3860858 Strep Assembly -- Assembly id#3860858
 ATCGAATTTGCCAACCAAGAAAAATATCCCTTGGATGGTTCTTGGCAATGCAAGCAATAT
 CATCGTTTCGTGATGGTGGGATTCGTGGATTTGTCATCTTGTGTGACAAGCTCAATAACGT
 TTCTGTTGATGGCTATACCATTGAAGCAGAAGCTGGGGCTAACTTGATTGAAACAACCTCG
 CATTGCCCTCCGTCATAGTTTAACTGGCTTTGAGTTTGCTTGTGGTATTCCAGGAAGCGT
 TGGCGGTGCTGTCTTTATGAATGCGGGTGCCTATGGTGGCGAGATTGCTCACATCTTGCA
 GTCTTGTAAGGTCTTGACCAAGGATGGAGAAATCGAAACCCTGTCTGCTAAAGACTTGGC
 TTTTGGTTACCGCCATTCAGCTATTCAGGAGTCTGGTGCAGTTGTCTTGTCAGTTAAATT
 TGCCCTAGCTCCAGGAACCCATCAGGTTATCAAGCAGGAAATGGACCGCTTGACGCACCT
 ACGTGAACCTCAAGCAACCTTTGGAATACCCATCTTGTGGCTCGGTCTTTAAGCGTCCAGT
 CGGGCATTTCGAGGTCAGTTTCGAATTTTCAGAAGCTGGCTTGAAAGGCTATCGTATCGGT
 GGCGTAGAAGTGTGAGAAAAGCATGCAGGATTTATGATCAATGTGCGCAGATGGAACGGCC
 AAAGACTACGAGGACTTGATCCAATCGGTTATCGAAAAAGTCAAGGAACACTCAGGTATT
 ACGCTTGAAAGAGAAGTCCGGATCTTGGGTGAAAGCCTATCGGTAGCGAAGATGTATGCA
 GGTGGTTTTTACTCCCTGCAAGAGGTAGTGGGGACCTGACAGAGCCCCGATCGGTTAATCT
 ATGAAAAAGAAGGAATTT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 610 | 807 | F | 66 aa |

> 3860858-1 ORF translation from 610-807, direction F
 VSEKHAGFMINVADGTAKDYEDLIQSVIEKVKEHSGITLEREVRIILGESLSVAKMYAGGF
 TPCKR*

Description:

unknown

Assembly ID: 3860890
 Assembly Length: 980bp

```
> 3860890 Strep Assembly -- Assembly id#3860890
CTGAAAAAACAGGTTTTGACTATGNAGATTGACAGACGACCGTTTCGGAGGTGCAGATATT
GATGCAGCAGGACCTCCCTTACCTGATGAAACCCTTAAGGCAAGTAGGGAAGCAGATGCT
ATCCTACTAGTAGCTATCGGTAGTCCCTCAGTATGATGGAGTAGCGGTTTCGCCCTGAACAA
GGCCTGATGGCTCTCCGTAAGAACTCAATCTTTACGCTAATATTCGTCTCTGTAAAAATCT
TTGACAGTCTCAAGTATTTGTCACCACTCAAACCGGAACGAATTTCTGGTGTAGACTTCG
TCGTGGTGCGTGAATTGACTAGGCGAGATTTACTTTGGAGATCATATCCTTGAAGAGCGC
AAAGCGCGTGATATCAACGACTATAGCTATGAGGAAGTGGAGCGGATTATTCGCAAAGCC
TTTGCCATCGAATTGCAAGAAATCGCAGAAAAATCGTTACTAGTATCGATAAGCAAAATG
TTCTAGCGACCTCAAACTCTGGCGGAAAGTAGCTGAGGAAGTCGCACAGGATTTCTCAG
ATGTAACCTTTGGAACACCAGCTGGTAGACTCAGCTGCTATGCTTATGATTACCAATCCTG
CTAAGTTTGATGTTATTGTAACGGAGAATCTTTTTGGAGATATTTTATCTGATGAATCAA
GCGTCTTATCTGGTACACTTGGGGTTATGCCATCAGCCAGTCATTCTGAAAATGGACCAA
GTCTCTATGAACCTATTCACGGTTCAGCACCTGATATTGCAGGTCAAGGAATTGCCAATC
CTATTTCCATGATTTTATCAGTTGTCATGATGTTGAGAGATAGTTTTCGGACGTTATGAGG
ATACAGAGCGTATCAAACGTGCTGTTGAGACAAGTCTGGCGGCAGGAATTTTAACGAGAG
ATATAGGAGGTGAGGCTTCAACAAAGGAAATGATGGAAGCTATTATTGCAAGGTTATGAA
GTTAGACGAAAAAATTCGAT
```

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 397 | 486 | F | 30 aa |

```
> 3860890-2 ORF translation from 397-486, direction F
VERIIRKAFAIELQEIAEKSLVLSISKMF*
```

Description:

unknown

Assembly ID: 3860952
 Assembly Length: 874bp

```
> 3860952 Strep Assembly -- Assembly id#3860952
TCGATCTAGAGAATTGCTCCAGAGCTTCCTGACCGTCCGCTGCCTCAATAGTTTCATAGC
CACAATCCGTCAAATAATCACTGACCCCTCACGGATCATCTCTTCATCTTCTACAATTA
AAATTTTCATACTTTAACTGCTCTCTATTTTTTATTTTTCTTAGAATAAATACCTACTCT
ATTTTCTATTATAGTCTCTTGCTGGCCTTTTGTATGTAAGCAACTGACCACTAGATAAAA
CGTTGTGAAATTCCTTCTCATAAATTCATAACTTTAGTATATTATTTAAGCACTAA
AGTACAAAGAAAGCAACTGAAAGCAATGATTTTCACCACTGCTTTCAGATTTATTTTGAA
```

TTGTTAAATAGCTATTCCTATCCACTATTCTTGAATAGAAACACAAGATGCAATCTTTAT
 TCCAGACTCATTTTTTTAAAAAATCAAATTTATTTCACCATCCAGCAAGAGCTCTTTTGGTT
 GTTTTCTAAGGAGATTGCTTGAAGCAAGCGCCATAACGAGAACCCTAGAACCAAGGCAA
 GGACAAAAATGATGATAAAGTCTGATGTCTGAATGGAAATGTCTAGGCTCGACAAGGTCT
 TGCTAAAGCCATCTACTTCTGCACCGCCACCAAGGTTAGAGGCTTGAGCCGCCTTACTAG
 CCTGTTTGGCAACACCTGAAGTCACATTGGCAAGGACAGTGTTCCTCAATTCGCACGGGCA
 GTGTAATTAGCTAGGAAGTAAGCANAACTAGAGCAGGGATAGCAATCAAGATAGATTTCG
 GTGATGAATTGACCCAAGATACTTGCTTGCTTGAGACCAATAGAGAGGAGGATTCCCCT
 TCCTTGCCGACGGGCATTGATCCAAAGACTGAGC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 449 | 715 | R | 89 aa |

> 3860952-1 ORF translation from 449-715, direction R
 VRIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQTSDFIIFVLA
 LVLVVLVLMALASSNLLRKQPKELLLDGE*

Description:

unknown

Assembly ID: 3860962

Assembly Length: 762bp

> 3860962 Strep Assembly -- Assembly id#3860962

CTTGTAACGGTCATAAAGTTTCTGCAAACTACCATCCTTGCTCCATTTAGTAACCAAGTT
 ATCAAGATAGTCGTTGAGCTCTGTATTTGATTTCTTGTTAACAATACCGTAGTCAGATGG
 CTTGAAACTATCATCTAGTAGTTCTGTGCGTTTAACTAGTGTAGCCAGATAGAATAGAGC
 GGTCAACGGAAAAGGCATCGATACGATGAGCGTGAAGGGAAGTAATCAATTCGGGTAGG
 AACCAAGTTCGACGAATTTAACTTCAGACCTTTCTTTTTACCCAGTTCAGTAATCAGGC
 GTTGGGTGATAGAACCTTGGGCGACTCCGATGGTTTTGCCGTTTAGGTCCCTCAATCTTTT
 TGATTTTGGCAGATTTATTGACCAAAAATCCAGAAGCGTCTGTGTAGTAGGGACTGGTAA
 AGTTGTAGAGTTTTTTGCGTTTCGTCCGTGATGGTAAAGGTCGCGATATCCATATCGACCT
 GTTCATTGTCTAGAAGGGGGCCGCGGTTTGTGCTGTAACCGGCACATAGTGAATCTTGA
 CCTTGAGTTCATCAGCTACCATTTTGGCCAAGTCGGTTTCGATACCAGAATAAGTACCGG
 TCTTGGGATCTTTGTTAACCAAAAATTGGGAACGTCTTGTGTTGACACCCGACAACCAAGTTC
 GCCTCTTTTTTGAATGTCTGCGATACTAGTATTAGCCTGGACTGGTTTGGCAGCAACAAG
 GCCGAAAAGGCTAATCAATAATGCTGATAAAAAGAATTTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| | | | | |

1 152 646 R 165 aa

> 3860962-1 ORF translation from 152-646, direction R
VSNKTFPILVNKDPKTGTYSGIETDLAKMVADELKVKIHYVPVTAQTRGPLLDNEQVDM
IATFTITDERKKLYNFTSPYYTDASGFLVNKSAKIKKIEDLNGKTIGVAQGSITQRLITE
LGKKKGLKFKFVELGSYPELITSLHAHRIDAFSVDRSILSGYTS*

Description:

cell adhesion factor PEB1 precursor - *Campylobacter jejuni*

Assembly ID: 3861268

Assembly Length: 1942bp

> 3861268 Strep Assembly -- Assembly id#3861268

CTCGAATTTTTGGTGCTCCAGAAACGGTTCCAGCAGGAAGCGTTGCTTTCAAGGCATCCA
TGGCAGTGAGTTCTGCAAGCAAACGTCCCTTGACCACACTGGTCAAATGCATGACGTAGC
GGAAGAGCTCCACCTCCATATACTTAGTAACCTGGACACTGGCCGTTTCAGAGATGCGGC
CAATATCGTTACGCCCCAAGTCTACCAACATTCGATGTTCTGCTGTTTCCTTCTCATCAG
AGAGGAGGTCAGTCGCCAAGGCCTTGCTTCTCCATCCGTAGCCCCCTCTTGGTCGCGTCC
CTGCAATCGGATTGGTTGTCACGATGCCATTTTTGACAGAAACCAAACCTTCTGGACTAG
CTCCGATGATTTGATAATCCCCAAAATCATACAAATAAAGGTAATTAGATGGATTAGTCA
CGCGGAGATTTCTGTAGAAGTCAAATGGATTTCCAGTTAACTTCTGCGTGAAGAAAACGC
TGGCTGAGTTACACATCGGAACATATCTCCGTTACGAATCAAGTCACGAGCTGTTTCTAC
CATTCCCTCAAACCTTATGTGGAGCGATATGCGGTTTGAAGTCAAGTGGTGATAAATCCAA
GTCTTCAAATTCATTTGGAGCAGGAATGCGTAATTCCTCAAGCACTTGGTTCAAGGATTT
TTCCAAGGCCTCTTGACTGCGCTCACTATAAAGTGCATCCTCTATGACATGTTATCTTCT
CCTTCTTGTGGTCAAAGACCATATAGCTCTCATAGACAAAGAAATGCATGTCGGGCGTC
CCAATTGTATCCTCAGGGATTTGACCAATTTCTTCATAAAGCGAAATCATATCGTAACCA
ACAAAACCAATGGCTCCCCCACCAAAAGGGAGGTCTGAATGGTGCTGGCTCTTATGAATC
ACTTCATAAAGGAAATCCAAGGGATCCCGATCAATCGCTTGACCATTTTGATAGAGAACT
CCATTTTCAAACCTTAATCTCAAAAACCTGGATTATAGGCTAGGATAGAAAAACGAGCTGTT
TCCTTGTCTCTCGGAATACTCTCTAAAATAACCTTATGTTGCCCTTTAAGCGCATATAA
GCCAAGATTGGTGATAAGACATCTCCATGAATGATTCGTTCCATTGTCATTTCCCTTTCA
GTTCTAATTCGAGTTTCGTGGCGACTGTATGAAAAATCCCCACGCAAAATAACTTGCGTGA
GGACGAAATTCGCGGTGCCACCTCAATTATAGGATTTCTCCTATCTCTCATTCCTGTCTC
AGATATCTCCTGTAACAGGCTGTGCGATAAAGGGCACTCCCTTGAGAATGATGTTTTCTT
CTCTCGTTTTAGATGAACCAACTTTACAGCTTTCTCTGCTTGTTTTTCAGCAACCACAAG
CTCTCTGTGAGAGAAAAGACTGTAATTTTTCCATCTATTATTTTCTAGCTTCTAGTAATC
TGCAATCGCAGCTAGGTCCCTTGCCCTCCACGACCAGAGACATTGATGAAGAGATGTTTCATC
TCGGTACACCTTTATACTCTTCGAAAATCTCTTCAAACCGCGTCAACGTGCGCTTGCCGT
AGGTATGGTTACTGACTTCGTCAAGTCTATCTGCAACCTCAAACAGTGTTTTGAGCTGA
CTTCGTCAAGTCTTATCGACAACCTCAAACAGTGTTTTGAGCAGCCTGCAGCTAGTTTCC
TAGTTTGCTCTTTGATTTTCATTGAGTATTATTTTCATTTTCTCCTGCAATTGAATTCTTG

67

 1 627 824 R 66 aa

> 3861270-1 ORF translation from 627-824, direction R.
 VSSSFFTPLKQLSKFLIIMAMSAIGLKTNLVAMVKSSGKSIVLGAVCWIAIILTSLGMQT
 LIGIF*

Description:
 unknown

Assembly ID: 3861288
 Assembly Length: 1571bp

> 3861288 Strep Assembly -- Assembly id#3861288
 AGAGCTGGTAATATTCCCAAAGAAACGGCTCAAATCGAATTAGAAAGCCTTCTGCAAAAA
 GGAATCCAGTCGCTCTGGTATCACGATGCTTTAACGGTATTGCCGAGCCTGTTTATGCC
 TACCAGGGTGGGGCGTACAGTTGCAAAAAGCAGGCGTTTCTTTGTAAAGAACTCAAC
 GCCCAAAAGCCCGCTTGAAACTCCTCATCGCCCTCAATGCCGACTAACAGGACAGGCT
 TTGAAAGACTATATGGAAGGCTAATACTCTTCGAAAATCTCTGCAAAACCACGTCAGCGTC
 GCCTTACCGTATGTAGAGCACAAAATCAGGAAATCTTCTCGATTCCCTGATTTTTTCTAT
 TTACGTTTTCTGTTGAGCTACGTTCTGTCAAACCATGAGGTAAGAGAACTTCACGTTCT
 TCCAACCTCTTCCTTATGCATAATCTTGGTCAACATACGCATACTAATGGCACCAAGGTCA
 TAAAGAGGTGGGCAATCGTTGTCAAGTTTGGACGGGTAAAGCGTGAGATTTGTGAATCA
 TCACTAGTAATAATTCGATAATCTTCTGGCACAGAAACACCTTATCAGCCAAACCGTTCA
 AGACTCCTGCTGCCAACTCATCACCTGTCACAACTGCTGCAGTTGCATTTGATGAAATCA
 AACGCTCTGCTAAGGCGTAACCATCATCATAGCTATATTTAGATTCAAATACCAACCCT
 CACTATAAGCGATTCTGCTTTTTTCAAGGTTTCTTGTAGCCAACTAAACGAACCTTAC
 CATTGATGTCATCCACTAGCGGACCGCTAACGAAAGCAATACGCTCATTTTCTTTAGCAA
 GGTAACCTCACTGCATCAATTGTTGCTTGCTTATAGTCAATATTGACACTTGGCAACTGGT
 GCTCAACATCGACAGTTCCCTGCGAGAACAATCGGAGTACGTGAACGCGAAAATTCTGAGC
 GAATTTTATCTGTCAAGTGATAACCCATATAGATAATGCCATCTACCTGCTTTGAAAAGA
 GGGTATTGACAACAGAACTTCTTTCTCGTTATCTTCATCGCTATTAGCTAGGACAATAT
 TGTACTTGTACATTTCTGCAATATCATCAATCCCCTTAGCCAAACTCGAAAAATAACCAT
 TGGTAATATTTGGAATCACGACACCGACAGTGGTTGTCTTTTACTTGCAAGACCACGCG
 CAACTGCATTTGGACGATAATCCAAACGATCAATTACCTCTAGCACTTTTTTACGGGTAT
 TCTCTTTTACATTTTATTTGCCATTGACCACACGGCTGACCGTCGCCATGGGAAACACCT
 GCTTCACGAGCGACATCATAAATGGTTACTGTATCATCTGCATTTCATTCCTTTTCTGTC
 CTTTCTATCTCCACACATTCCTTTTACAAGTAGAAGTGCTGAATTGAAAGCTCTATATCTT
 ACTTACAAAAATGAAGATGTGAAAATTTTCGTTTTTCATATTTCTACTTATTCCATTCTATC
 ACTAATTGTAAACACTTTCAAGTGTTTTTTGAAGATTGATTGAAAAAATTTTCATAGAAAA
 CCTAGGTTTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 357 | 572 | R | 72 aa |

> 3861288-1 ORF translation from 357-572, direction R
 VPEDYRIITSDDSQISRFRPNLTTIAQPLYDLGAISMRMLTKIMHKEELEEREVLLPHG
 LTERSSTRKRK*

Description:

GLUCOSE-RESISTANCE AMYLASE REGULATOR. - BACILLUS SUBTILIS.

Assembly ID: 3861306

Assembly Length: 1682bp

> 3861306 Strep Assembly -- Assembly id#3861306

CTGACGTAAAAAGATTTTCGGAAAAGTATCATCATCTATTTTAGACCATTTTCTTATAA
 TAACCATTTTATTTTATTTTGTCAAGGTCTTTGAATTCTTTCTTAAACAAGCCTTGTAAT
 CTCTACTTTTGAAGAATTTATTTTTCCTTACTGACAAGATTTGAGACGGTAGGAATCATT
 GAAAATAACCTAGCCAACATCAATCACAATCATTTCTCCTTTCTCAATTACACTAAATTA
 TAGTGTATTGAATCTATAACAGTGCACCTTGGCTGCTAAAATATTTCTATAAATTAATTT
 GACTTTCCTGATAGAGTTGTTTACATCTTATTTCAATTCACATACTTTCCCTTATACTC
 AATGAAAATCAAAGCGCAAAGTAGGAAGCTAGCCACAGGCTGCTCAAAGCACTGCTTTGA
 GGTGTAGATAAGACTGACGAAGTCAGTTACATATATCTACGGCAAGGCGAAGCTGACGC
 GGTTTGAAGAGATTTTCGAAGAGTATAAAGTTTGTCTGTATCTTTTCAGAAAAATAAGG
 TATACTGTATGTAAACGATTTCAAAGGAGTCCAGTTATGGCAAAAACATTTTTTATTTCCA
 AATAAACAGAGCATTTTAGGAGAACAAAGAGATTTTGAATGCCAAGTCGATCTTGGCTATG
 ATGTAGTCTATCTCCGTCAGCCTCTTAATCGTCTCGAGTATATTGAGTGTGCGATAGTGG
 GGCAATCACAAATTTCTTTTAAAGGTCAGTTATGCTGATGGTCAAAGGCTTACCGTGTCTG
 ATCTTCCTGACCTACTAACAAAGACAGACTGGCAGATTATCAAGTCATTTTTTAGATGTTT
 TGCTTGCTTATACAGGGACTGATATTGAAGGGCTAGATGGTTTTGATTTTGAAGCTTATT
 TCCAAGCAAGTATTCAAGCCTATCTAGCAGACCCTGTAGCTCGTTTTACGATTTGCCAAC
 GAATTTTTAATCCTATTTTCTTTAGTCGTGAGAACTTGAAAAGCTTTTTAGAGGCAGATG
 GCTTGGCTCAGTTTGAAGCGCGTGTGCGTGCGGTTCAAGAGACAGATGCCTACTTTGCGA
 GAGTTTCCTTCTATCAGGATGGAGAAGGAAAAGTGCATGGCGTTTACCATCTAGCTCAAG
 GAGTCAAGACAGTTTTTACCGAGAGAACCGTTTTGTTCCCTGCAGCCTATATTGAGCGAATTG
 GTGGATAAGGAAGTCCAGTGGGAGATTGACTTGGTTCAAATCACAGGAGACGGCTCTAAA
 CCAGAAGACTATGAATCCATAGCTCGCTTGGACTATGCAAAATTTCTTAGAGGTATTACCC
 CCATCTTTTTTACCACCAACTAGACGCCAATCAAATAGAAATACAACCCATCCTAGGACAA
 GATTTTTAAAACATTAGCACAAAGAAAAGTAAAGCAGAAGCAGGTCAATCGACTTGCTTTTT
 TGACATAGAAAAAATCCTGCCAAGGATGACAGGATTGCTACTCAATGAAAATCAAAGAGC
 AAAGTAGGAAGCTAGCCGAGGCTGTACTTGAGTACGGTAAGGCGAAGCTGACGTGGTTT
 GAATTTGATTTTTCGAAGAGTATGAATTTTAAAGAAAGGCCAAGATACGAAGATAATCTCC
 AATCAGTGCCACTTCAGCTTCCAAGAAGAAGAAGATTATAACTCCCCTTCCCCAAGGACA

GA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 717 | 1208 | F | 164 aa |
| 2 | 1201 | 1410 | F | 70 aa |

> 3861306-1 ORF translation from 717-1208, direction F
 VGQSQFLFKVSYADGQKAYRVDLPDLLTKTDWQIIKSFLDVLLAYTGTDIEGLDGFDFEA
 YFQASIQAYLADPVARFTICQRIFNPIFFSRENLKSFLADGLAQFEARVRAVQETDAYF
 ARVSFYQDGEKGVHGVYHLAQGVKTVLPREPFVPAAYIERIGG*

Description:

unknown

> 3861306-2 ORF translation from 1201-1410, direction F
 VDKEVQWEIDLTVQITGDGSKPEDYESIARLDYAKFLEVLPPSFYHQLDANQIEIQIPILGO
 DFKTLAQEK*

Description:

unknown

Assembly ID: 3861334
 Assembly Length: 3041bp

> 3861334 Strep Assembly -- Assembly id#3861334
 ATCGAATTAAAAATGAGGTATTCAGGCTTGTGATTTTCTATGGAAGTTAATAGTGATTGC
 CTCTAATGCTTACAAGTGATATTAATAAGAGGACCTAGTGATGTCAATCATTTCAACT
 GATTTAACCCTTTTCAAATAGATGATACATTGAAAGCAGCCTTGCGAGAAGATGTTTCAT
 TCCGAAGATTACAGTACCAATGCCATTTTTGATCATCATGGCCAAGCCAAGGTGTCGCTT
 TTTGCCAAGGAAGCTGGTGTTTTAGCGGGGCTAACCCTTTTCAAAGGGTTTTTACCCTA
 TTTGATGCCGAGGTGACCTTCCAGAATCCTCATCAATTTAAGGATGGGGATCGTTTGA
 AGTGGCGATTTGGTTTTAGAAATCATAGGCTCGGTGAGAAGTCTCTTAACATGTGAACGC
 GTTGCCCTTGAATTTTTTACAACATTTATCAGGGATCGCTTCGATGACAGCTGCTTATGTA
 GAAGCCTTAGGCGATGATTGCATTAAGGTATTTGATACTCGAAAACTACTCCTAATTTA
 CGTCTTTTTGAGAAATATGCCGTGAGAGTTGGCGGTGGCTATAATCATCGCTTTAATTTA
 TCAGATGCTATCCTGCTAAAAGACAATCACATTGCGGCAGTAGGTAGTGTTCAAAGGGCA
 ATTGCTCAAGCGCGTGCTATGCTCCTTTTGTGAAAATGGTCGAGGTGGAAGTGGAAGC
 CTTGCTGCTGCCGAAGAAGCTGCGGCGGCGGGTGCTGATATTATCATGTTGGATAATATG
 TCATTGGAACAGATTGAACAGGCCATTACCCTAATTGCAGGACGTTCTCGGATTGAATGT
 TCTGGAAATATTGATATGACCACTATTAGCCGTTTTCTGTTGGTTAGCGATTGATTACGTC
 TCCAGTGGTAGTTTAACCCATAGTGCTAAGAGTCTTGATTTTTCCATGAAGGGTTTAACC

TACCTTGATGTCTAAGTTGTAAATAAACTAACTTTTTAAAGGATGTCTTTCCTCTAGAA
 CGAGTTTTATGTCAGATAGTTTTAAACGCCTCTTCAAATATAGTAAAATGAACCAAAAATA
 GTACACAATGTGGTATAATCTTCTTATGGCATATTCAATAGATTTTCGTAAAAAAGTTCT
 TTCTTATTGTGAGCGAACAGGTAGTATAACAGAAGCATCACACGTTTTCCAAATCTCACG
 TAATACCATTTATGGCTGGTTAAAGCTAAAAGAGAAAACAGGAGAGCTAAACCACCAAGT
 AAAAGGAACAAAACCAAGAAAAGTTGATAGAGATAGACTTAAAACTATCTTACTGACAA
 TCCAGACGCTTATTTGACTGAAATAGCTTCTGAATTTGGCTGTCATCCAACCTACCATCCA
 CTATGCGCTCAAAGCTATGGGCTACACTCGAAAAAAGGACCACACCTACTATGAACAAGA
 CCCAGAAAAAGTAGCCTTATTTCTTAAAAATTTTAATAGTTTAAAGCACCTAGCACCTGT
 TTAGATTGATGAAACAGGATTCGATACTTATTTTTATCGAGAATATGGTCGCTCATTTAA
 AGGTCAGTTAATAAGAGGTAAAGTATCTGGAAGAAGATATCAGAGGATTTCTTTGGTTGC
 AGGTCTAACAAATGGTGAGTTAATCGCTCCAATGACTTACGAAGAGACGATGACGAGCGA
 CTTTTTTGAAGCATGGTTTTCAGAAGTTTCTCTTACCAACATTAACCACACCATCGGTTAT
 TATTATGGATAATGCAAGATTCCATAGAATGGGTAAGTTAGAAGTTTTATGCGAGGAGTT
 TGGGCATAAACTTTTACCTCTTCCCTCCCTACTCGCCTGAGTACAATCTTATTGAGAAAAC
 ATGGGCTCATATCAAAAAGCACCTCAAAAAGGTATTACCAAGTTGCAATACCTTTTATGA
 GGCTCTTTTGTCTGCTCTTGTTCATTGACTATAGTTTACGGATACAGTTGGGAAAAGA
 AGTTAAATGTAGTTGGATTTCCACTAAAGGTGATGAGTAAGTTTTTGTATCTGAACCTG
 ATTGGCCGCAAGCAGCTAAAAGCAAAGCAGATGCAAAAGTCAGACCTGCACCAAGGACAC
 GCTTCTTTATGTTTCATCTTCTTCTCTTAATAGTGGGAATTTGTAAAGTTAATTGAATT
 TCAAGAATGAAGGTTTTATAAACTTTGGTTATAAAAAACAAAGGATTTCTGTCTTTTATA
 CAGTCCTCCCTTGTTTTTATACGATTTCAATTTTAAATTTTTCTGCAAAAAATATTTAT
 AGTAATTCACACAGAAAGCATCCCATGGAAGTAAGATTTGTTTTTCAAAGACTTCTTGA
 GCTAGGGTGTTTTCAATCAAGACAGATTTGACTTTTCTTCTACTGTCAAGTCTTGCTCT
 TCATTGGACAAGTTAGCCACAAGTAGGAAGCGACGGTCGCCATCCTTACGTATATAAGCA
 AAGACCTTATCAGCCGTATCAAGCAATTCAAAGTCAGCTCGAATTAGCCAACTATTCTCC
 TTGCGAATTTGGACCAGTTTCTGATAGGTATAGAAAATAGAATCTGGATTTGCCAGCGCT
 TCTTGGACGTTGATCATCTCGTAATTTGGATTAACTGCCAACCAAGGTTGACCTGTTGAG
 AAACCAGCGTTTTTGCTCTCGTCCCATTGCATAGGGGTACGGGCATTGTACGTCCAATA
 ACACGGATACTGTCCATGATTTCTTGCATCGGAACACCTTTTTCAAGAGCCTCACGCGCA
 TAGTTGAGAGATTCAATATCTTCTACTTGATCCAGTGTTCAAACGGATAGTTGGTTCATC
 CCAATCTCCTCACCTTGGTAGATATAAGGAGTTCCCTCTCATAAGATGAAGCAAGATTGCA
 AAGGCTTTGGCAGATTTTTCGCGGTATTCCTTGGTCATTTCCCCAGATTGAGACAATACGA
 GGGAGGTCATGGTTGTCCAGAAGAGGGAATTCCAGCCGTCCTCAACTCCTAACTCTGTC
 TGCCATTTGTTGAAGATTTCTTTTAACTTAGCGATATTCAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 76 | 975 | F | 300 aa |

> 3861334-1 ORF translation from 76-975, direction F
 VILKIEDLVMSIISTDLTPFQIDDTLKAALREDVHSEDYSTNAIFDHHGQAKVSLFAKEA

GVLAGLTVFQRVFTLFDAEVTFQNP HQFKDGDRLTSGDLVLEIIGSVRSLLTCERVALNF
LQHLSGIASMTAAYVEALGDDCIKVFDTKRKTPNLRLFEKYAVRVGGGYNHRFNLSDAIL
LKDNHIAAVGSVQRAIAQARAYAPFVKMVEVEVESLAAAEAAAAGADIIMLDNMSLEQI
EQAITLIAGRSRIECSGNIDMTTISRFRGLAIDYVSSGSLTHSAKSLDFSMKGLTYLDV*

Description:

PROBABLE NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING)
(EC 2.4.2.19) (QUINOLINATE PHOSPHORIBOSYLTRANSFERASE
(DECARBOXYLATING)) (QAPRTASE) (FRAGMENT). - BACILLUS
SUBTILIS (BLAST)

Assembly ID: 3864148

Assembly Length: 4694bp

> 3864148 Strep Assembly -- Assembly id#3864148

TTAATTTAAATTCTTAAAATTTTTTCATAATAATCTCCCTATAAAAAATAAGTCGCCCAA
TCAGGCGGCTTATTTTTTTGAAAAATGGGCTTGGTGCCGAGAATAAATAGCTTAGTGAT
AGAAGAAAATGGGGAAATATGGTATAATGAAACGATAGATTTTTGAATAGGAATAAGATC
ATGTTTGGATTTTTTAAGAAAGATAAAGGCTGTGGAAGTAGAGGTTCCGACACAGGTTCC
TGCTCATATCGGCATCATCATGGATGGCAATGGCCGTTGGGCTAAAAAACGTATGCAACC
GCGAGTTTTTGGACATAAGGCGGGCATGGAAGCATTGCAAACCGTGACCAAGGCAGCCAA
CAAACCTGGGCGTCAAGGTTATTACGGTCTATGCTTTTTCTACGGAAAACCTGGACCCGTC
AGATCAGGAAGTCAAGTTTATCATGAACTTGCCAGTAGAGTTTTATGATAATTATGTCCC
GGAACCTACATGCGAATAATGTTAAGATTCAAATGATTGGGGAGACAGACCGCCTGCCTAA
GCAAACCTTCGAAGCTTTAACCAAGGCTGAGGAATTGACTAAGAACAACACAGGATTGAT
TCTTAATTTTGCTCTTAACCTATGGTGGACGTGCTGAGATTACACAGGCGCTTAAGTTGAT
TTCCCAGGATGTTTTAGATGCCAAAATCAACCCAGGTGACATCACAGAGGAATTGATTGG
TAACCTATCTCTTTACCCAGCATTTGCCTAAGGACTTACGAGACCCAGACTTGATTATCCG
TACTAGTGGAGAATTGCGTTTGAGCAATTTCCCTTCCATGGCAGGGAGCCTATAGTGAGCT
TTATTTTACGGACACCTTATGGCCTGATTTTGACGAAGCGGCCTTGCAGGAAGCTATTCT
TGCCTATAATCGTCGCCATCGCCGATTTGGAGGAGTTTAGGAGGAAAATATGACCCAGGAT
TTACAGAAAAGAACCTTGTTATGCAGGGATTGCCCTGACTATTTTCCTACCAATTTTAAT
GATTGGGGGCTCTTGCTTCAGATAGCAATCGGAATCATANCCATGCTAGCCATGCATGAA
CTTTTGAAGATGAGAGGTCTAGAGACCATGACGATGGAGGCCTCTTGACCTCTTTGCAC
NTTNGTATTGACCATTCCCCTGGAATCGAATTACCTGACTTTTTTGGCAGTTGATGGGAA
TGTGGTTGCCTATAGTGTTTTGATTTCAATCATGTAGGAACGACCGTTTTTAGCAAGTC
TTATACGATTGAGGATGCGGTTTTCCCTCTTGCTATGAGCTTCTACGTGGGCTTTGGATT
TAATGCTTTACTAGATGCTCGTGTGTCAGGTTTGGACAAGGCTCTCTTAGCCTTGTGTAT
CGTCTGGGCGACAGACAGTGGTGCCTATCTTGTTGGGATGAACTATGGGAAACGAAAGTT
AGCACCAAGGGTATCGCCTAATAAAAACCTTGAGGGTGCCTTGGGTGGTATTTTAGGAGC
AATTTTAGTAACCATTATCTTTATGATAGTTGACAGTACAGTTGCTCTTCCATATGGAAT
TTACAAGATGTCAGTCTTTGCTATTTTCTTTAGCATTGCTGGACAATTTGGTGATTACT
AGAAAGTTCGATCAAACGTCATTTTGGTGTTAAGGATTCTGGGAAATTTATCCCTGGACA

TGGTGGTGT TTTTGGATCGTTTCGATAGTATGTTGCTTGTATTTCCAATCATGCACTTATT
TGGACTCTTTTAATCAAAAGACGGAGGAAACGCTATGCTCGGAATTTTAACCTTTATTCT
GGTTTTTGGGATTATTGTAGTGGTGCACGAGTTCGGGCACTTCTACTTTGCCAAGAAATC
AGGGATTTTAGTACGTGAATTTGCCATCGGTATGGGACCTAAAATCTTTGCTCACATTGG
CAAGGATGGAACGGCCTATACCATTCGAATCTTGCCCTCTGGGTGGCTATGTCCGCATGGC
CGGTTGGGGTGATGATACAACTGAAATCAAGACAGGAACGCCGTGTAGTTTGACACTTGC
TGATGATGGTAAGGTTAAACGCATCAATCTCTCAGGTAAAAAATTGGATCAAACAGCCCT
CCCTATGCAGGTGACCCAGTTTGATTTTGAAGACAAGCTCTTTATCAAAGGATTGGTTCT
GGAAGAAGAAAAAACATTTGCAGTGGATCACGATGCAACGGTTGTGGAAGCAGATGGTAC
TGAGGTTCCGATTGCACCTTTAGATGTTCAATATCAAATGCGACTTTATCTGGGGCAA
CTGATTACCAATTTTGCAGGTCCATGAACAATTTTATCTTAGGTGTTGTTGTTTTTGG
GTTTTAATCTTTATGCAGGGTGGTGTGTCAGAGATGTTGATACCAATCAGTTCCATATCATG
CCCCAAGGTGCCTTGGCCAAGGTAGGAGTACCAGAAACGGCACAAATTACCAAGATTGGC
TCACATGAGGTTAGCAACTGGGAAAGCTTGATCCAAGCTGTGGAAACAGAAACCAAAGAT
AAGACGGCACCGACTTTGGATGTGACTATTTCTGAAAAGGGAGTGACAAACAAGTCACT
GTTACACCCGAAGATAGTCAAGGTCGTTACCTTCTAGGTGTTCAACCGGGGGTTAAGTCA
GATTTTCTATCCATGTTTGTAGGTGGTTTTACAACCTGCTGCTGACTCAGCTCTCCGAATT
CTCTCAGCTCTGAAAATCTGATTTTCCAACCGGATTTGAACAAGTTGGGTGGACCTGTT
GCTATCTTTAAGGCAAGTAGTGATGCTGCTAAAAATGGAATTGAGAATATTCTTGTA
CTTGGAATGATTTCCATCAATATTGGGATTTTTAATCTTATTCCGATTCCAGCCTTGG
TGGTGGTAAGATTGTGCTCAATATCCTAGAAGCCATCCGCCGCAAACCATTGAAACAAGA
AATTGAAACCTATGTCACCTTGGCCGGAGTGGTCATCATGGTTGTCTTGATGATTGCTGT
GACTTGGAATGACATTATGCGACTCTTTTTTTAGATAATCGAGGAATATTATGAAACAAAG
TAAATGCCTATCCCAACGCTTCGCGAAATGCCAAGCGATGCTCAAGTTATCAGCCATGC
TCTTATGTTGCGTGCTGGTTATGTTGCGCAAGTTTCAGCAGGTGTTTATTCTTATCTACC
ACTTGCCAACCGTGTGATTGAAAAAGCTAAAAACATCATGCGCCAAGAATTGAAAAGAT
TGGTGTGTTGAGATGTTGGCTCCAGCCCTTCTTAGTGCAGAATTGTGGCGTGAATCAGG
TCGTTACGAAACCTATGGTGAAGACCTTTACAACTGAAAAACCGTGAAAAATCAGACTT
TATCTTAGGTCCAACCTCACGAAGAAACCTTTACAGCTATTGTCCGTGATTCTGTTAAATC
TTACAAGCAATTGCCACTCAACCTTTATCAAATTCAGCCCAAGTATCGTGATGAAAAACG
CCCACGTAATGGACTTCTTCGTACACGTGAGTTTATCATGAAGGATGCTTATAGTTTCCA
CGCTAACTATGATAGTTTGGATAGTGTATGATGAGTACAAAGCAGCCTATGAGCGTAT
TTTCACTCGTAGTGGTTTAGACTTCAAGGCTATTATTGGTGACGGTGGAGCCATGGGTGG
TAAGGATAGCCAAGAATTTATGGCCATTACATCTGCTCGTACAGACCTTGACCGCTGGGT
TGTCTTGGACAAGTCAGTTGCCTCATTTGACGAAATTCCTGCAGAAGTGCAAGAAGAAAT
CAAGGCAGAATTGCTCAAATGGATAGTCTCTGGTGAAGATACCATTGCTTACTCAAGTGA
GTCTAGCTATGCAGCTAACTTAGAAATGGCAACAAACGAGTACAAACCAAGCAACCGTGT
TGTCGCTGAAGAAGAAGTTACTCGTGTTGAAACGCCAGATGTTAAATCAATTGATGAAGT
TGCAGCCTTCCTCAATGTTCCAGAAGAACAAACGATTAAAAACCCTCTTCTACATTGCAGA
TGGTGAGCTTGTTGCAGCCCTTCTAGTTGGAAATGACCAACTCAACGAAGTCAAGTTGAA
AAATCACTTGGGAGCAAATTTCTTTGACGTTGCTAGCGAAGAAGAAGTGGCGAATGTTGT
TCAAGCAGGATTTGGTTCACTTGGACCAGTTGGTTTGCCAGAGAATATTAAATATTATGC
AGATCGTAAGGTGCAAGATGTTTCGCAATGCAGTTGTGCGGTGCTAACGAAGATGGCTACCA

CTTGACTGGTGTGAACCCAGGCCGTGATTTTACTGCAGAATATGTGGATATCCGTGAAGT
 TCGTGAGGGGTGAAATTTCCCCAGATGGACAAGGTGTCCCTTAACCTTTGCGCGTGGTATTGA
 GATCGGTCATATTTTCAAACCTCGGAACCTCGCTATTTCAGCAAGCATGGGAGCAGATGTCTT
 GGATGAAAATGGTCGTGCTGTGCCAATCATCATGGGATGTTACGGTATCGGTGTCAGCCG
 TCTTCTTTTACGACGTGATGGAGCAACACGCTCGCCTCTTTGTTAACAAAACGCCAAAAGG
 TGAATACCGTTACGCTTGGGGAATCAATTTCCCTAAAGAATTGGCACCATTGTGATGTGCA
 TTTGATTACTGTTAATGTCAAGGATGAAGAAGCGCAAGCCTTGACAGAAAACTTGAAGC
 AAGCTTGATGGGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 212 | 940 | F | 243 aa |
| 2 | 1202 | 1753 | F | 184 aa |
| 3 | 2750 | 3037 | F | 96 aa |

> 3864148-1 ORF translation from 212-940, direction F

VEVEVPTQVPAHIGIIMDGNGRWAKKRMQPRVFGHKAGMEALQTVTKAANKLGVKVITVY
 AFSTENWTRPDQEVKFIMNLPVEFYDNYVPELHANNVKIQMIGETDRLPKQTFEALTKAE
 ELTKNNTGLILNFALNYGGRAEITQALKLISQDVLDKINPGDITEELIGNYLFTQHLPK
 DLRDPDLIIRTSSELRLSNFLPWQGAYSELYFTDTLWPDFDEAALQEAILAYNRRHRRFG
 GV*

Description:

unknown

> 3864148-2 ORF translation from 1202-1753, direction F

VVAYSVLISIMLGTTVFSKSYTIEDAVFPLAMSFYVGFGFNALLDARVAGLDKALLALCI
 VWATDSGAYLVGMNYGKRKLAPRVSPNKTLEGALGGILGAILVTIIFMIVDSTVALPYGI
 YKMSVFAIFFSIAGQFGDLLESSIKRHFGVKDSGKFIPGHGGVLDRFDSMLLVFPIMHLF
 GLF*

Description:

CDP-diglyceride synthetase (cdsA) homolog - Haemophilus
 influenzae (strain Rd K W20)

> 3864148-10 ORF translation from 2750-3037, direction F

VDLLLSLRQVVMLLKME LRIFLYFLAMISINIGIFNLIPALDGGKIVLNILEAIRRKP
 LKQEIETYVTLAGVVIMVVLMI AVTWNDIMRLFFR*

Description:

unknown

Assembly ID: 3864172
 Assembly Length: 1352bp

> 3864172 Strep Assembly -- Assembly id#3864172

CTCGTAAGTTCGGAAGCTATCTACACAAGAAATTAACCGCTGCCTAAAGGAGAAGCCATG
 TCAACATATAACTGGGATGAGAAGCATATCCTTACCTTTCCTGAAGAAAAAGTAGCCCTT
 TCTACTAAGGATGTCCATGTTTACTATGGTAAAAATGAATCCATTAAGGGGATTGATATG
 CAATTTGAAAGAAATAAAATTACAGCTTTGATTGGTCCGTCGGGATCGGGGAAATCTACC
 TACTTACGCAGTCTCAATCGCATGAATGATACCATTGATATTGCTAAAGTAAGTGGGCAG
 ATTCTCTATCGTGGAATTGATGTCAACCGTCCAGAAATCAACGTTTATGAAATGCGTAAA
 CACATTGGAATGGTTTTTCAACGCCCAATCCATTTGCTAAATCGAATTTACCGTAATAT
 TACCTTTGCGCATGAACGTGCTGGAGTTAAGGATAAGCAAGTCCTAGATGAAATCGTAGA
 AACCTCCCTTAGTCAGGCTGCCCTTTGGGATCAGGTTAAAGACGATCTCCACAAGTCAGC
 CTTGACCTTATCAGGTGGTCAGCAACAACGTCTCTGTATCGCTCGTGCCATCTCTGTAA
 GCCAGATATCCTCTTAATGGATGAGCCAGCCTCAGCCTTGGATCCGATTGCGACCATGCA
 ACTAGAAGAGACCATGTTTGGAGCTCAAGAAAACTTTACCATCATCATTTGTAACGCATAA
 TATGCAGCAGGCTGCTCGTGCAAGTGACTATACAGGCTTCTTTTACTTGGGTGATTTGAT
 TGAGTATGACAAGACTGCAACTATTTTCCAAAATGCCAAGCTACAGTCCACCAATGACTA
 TGTATCTGGTCACTTTGGTTAGAAAGGAAACCGTATGACAGATGCGATTTTACAGGTATC
 AGACCTGTCCGTTTATTATAATAAAAAGAAGGCTTTGAATAGTGTTTCCCTATCTTTCCA
 ACCTAAGGAAATTACAGCCTTGATTGGTCCATCTGGATCAGGGAAGTCAACCCTCCTCAA
 GTCTCTCAACCGCATGGGAGATCTCAATCCAGAGGTGACCACAACCTGGATCCGTGGTGTA
 CAATGGTCACAACATCTACAGTCCGCGTACAGATACGGTTGAATTACGTAAGGAAATCGG
 AATGGTTTTTCCAACAACCTAATCCTTTCCCTATGACTATCTATGAGAATGTTGTCTACGG
 GCTTCGTATCAATGGAATTAAGGATAAGCAGGTTCTGGATGAAGCCGTAGAAAAAGCCTT
 GCAAGGTGCCTCTATCTGGGATGAGGTCAAGGATCGTCTATATGATTCAGCTATTGGATT
 GTCAGGTGGTCAACAGCAGCGTGTCTGCGTGG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 311 | 862 | F | 184 aa |

> 3864172-2 ORF translation from 311-862, direction F

VELMSTVQKSTFMKCVNTLEWFFNAPIHLLNRIYRNITFAHERAGVKDKQVLDEIVETSL
 SQAALWDQVKDDLHKSLTSLGGQQRLCIARAI SVKPDILLMDEPASALDPIATMQLEE
 TMFELKKNFTIIIVTHNMQQAARASDYTGFFYLGDLIEYDKTATIFQNAKLQSTNDYVSG
 HFG*

Description:

HYPOTHETICAL ABC TRANSPORTER (ORF75). - BACILLUS SUBTILIS.
 (BLAST)

Assembly ID: 3864180
Assembly Length: 2258bp

```
> 3864180 Strep Assembly -- Assembly id#3864180
AACTTCGACCGTGATAAACAAGCTGAGCTTTGACATACTTGTAGCCAACCTAAAAGCCGT
TCTTCAAGGCCTCAAACCAGCTGCAACTCATTGAGGAAGCCTGGATGAAAATGAAGTGGC
TGCCAAATGTTGAAACCAGACCAGAACTCATCACAAGAACTGAAGAAATTCATTTGAAGT
TATCAAGAAAGAAAATCCTAATCCCAGCTGGTCAGGAAATATTATCACAGCAGGAGTCAA
AGGTGAACGAACTCATTACATCTCTGTACTCACTGAAAATGGAAAAACAACAGAAACAGT
CCTTGATAGCCAGGTAACCAAAGAAGTTATAAACCAGTGGTTGAAGTTGGCGCTCCTGT
AACTCACAAGGGTGATGAAAGTGGTCTTGCACCAACTACTGAGGTAAAACCTAGACTGGA
TATCCAAGAAGAAGAAATTCATTTACCACAGTGAAGTGGTAAAATCCACTCTTACTCAA
AGGAAAAACACAAGTCATTACTAAGGGTGTCAATGGACATCGTAGCAACTTCTACTCTGT
GAGCACTTCTGCCGATGGTAAGGAAGTGAACACACTTGTAATAGTGTCTGACACAGGA
AGCCGTTACTCAAATAGTCGAAGTCGGAAGTATGGTAACACATGTAGGCGATGAAAACGG
ACAAGCCGCTATTGCTGAAGAAAAACCAAACTAGAAATCCTAAGCCAACAGCTCCTGC
TGAGGAAAGCAAAGCTCTTCCTCAAGATCCAGCTCCTGTGGTAATAGAGAAAAAACTTCC
TGAAACAGGAACTCACGATTCTGCAGGGACTAGTAGTCGCAGGACTCATGGCCACACTAG
CAGCCTATGGACTCACTAAAAGAAAAGAAGACTAAGTCTTTTCGATAAAAAATAAACAGC
GAGATTGAAGCTCGCTGTTTATTTTTTAATTAATCACCTAGTCCAAGACGTTCAAAGATA
TCATCCACTCGTTTGGTGTAATAAACTGGGTTGAAGATTTTCATCGATTTCTTCTGTGTG
AGACGTGATGTTACTTCTGAATCTGCCTCAAGAAGTGGTTTAAAGTCTACTTGGTTGTCC
CAAGAGTAGGCTGTTTTTGGTTGCACCAAGTCATAGGCTTGCTCACGGGTCATGCCTTTT
TCAATCAATGTCAACATAGCCCCGTTGGCTAAAGATAAGACCAAAAGTCGAGTTCATGTTT
CGGATCATATTTTCTGGGAAGACTGTCAAGTCTTTGACGATATTTCCAAAACGGTTGAGC
ATGTAGTCAATCAAATGGTCGTATCTGGTGTGATGATACGCTCAGCTGATGAGTGAGAA
ATATCGCGTTTCGTGCCAGAGAGCGACGTTTTTCATAAGCCGTAATCATGTGACCACGAATG
ACACGCGCCAGACCAGTCATATTTTCAGAACCGATTGGGTTGCGTTTGTGAGGCATTGCT
GAAGACCCTTTTGGCCCTTTAGCAAAGAACTCTTCTACTTCGCGTTGCTCAGATTTTGT
AGACCACGAATCTCAGTCGCCATACGTTTCGATTGAAGTCGCAATGCTGGCAAGAACCGCA
AAGTACTCAGCGTGAAGGTCACGAGGAAGGACTTGTGTTAAAGATTTCCTGGGCACGGAT
GCCAAGATTTATCGCAGACATACTCCTCTACAAATGGTGGGATATTGGCAAAGTTCCCAA
CCGCACCAGAAATCTTACCAGCTTCTACACCAGCAGCCGCATGCTCGAAGCGCTCGATAT
TGCGTTTCATTTTCGCTGTACCAAGTTGCTAATTTAAGACCAAAGGTTGTGGGCTCAGCGT
GCACACCATGAGTACGCCCCATCATGATGGTGAACCTGTGCTCCTTGGCCTTGTACGCGA
TGATATTAGTGAAGTTTTCAAGGTCACGACGGATGATGTCGTTGGCCTGCTTGTAGAGGT
AACCATAAGCAGTATCCACCACGTCGGTAGAAGTTAACCCTAGTGAACCCACTTGCGCT
CTTCACCAAGAGTCTCAGAAACCGCACGCGTGAAAGCCACCACATCGTGGCGCGTCTCCT
GCTCAATTTCCAAAATACGGTCGATGTCAAAGTCCGCTTCTTGCGAATCAAAGCCACAT
CTTCCTTAGGGATTTCCCCCACTCAGCCCATGCCTCGTCAGAGAGGATTTCCACCTCAA
GCCAAGCACGGTATTTATTTTCTTCACTCCAAATATTGCGCATCTCAGGGCGAGAGTAAC
GGTTGATCATGTGTTAATTTTCTTCTTCTTAAGAT
```

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 930 | 1616 | R | 229 aa |

> 3864180-2 ORF translation from 930-1616, direction R
 VPKESLTQVLPRDLHAIEYFAVLASIATSIERMATEIRGLQKSEQREVEEFFAKGQKGSSA
 MPHKRNPISSENMTGLARVIRGHMITAYENVALWHERDISHSSAERIITPDTTILIDYML
 NRFGNIVKNLTVFPENMIRNMNSTFGLIFSQRAMLTLIEKGMTREQAYDLVQPKTAYSWD
 NQVDFKPLLEADSEVTSRLTQEEIDEIFNPVYYTKRVDDIFERLGLGD*

Description:

ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL). -
 BACILLUS SUBTIL IS.

Assembly ID: 3864184

Assembly Length: 4392bp

> 3864184 Strep Assembly -- Assembly id#3864184

CCCTTTTGCCCTCTCCCTTTGGTGCAGATTCTTTTGGGAATTGTGATTGGTCTCTTTTAC
 CCAATACTGACTTTCATCTTAATACGGAGTTGTTTTGGCCTGGTTATCGGACCCTTGCT
 TTTCCGAGAGGCTGAAGAAGCAGATGTTACGGCTATTTTAAAACACTGGCGAATCATTGT
 TTATCTCATATTTCCAGTGATTTTTATCTCGACCCTGAGTTTGGGTGGCTTGGCCCATCT
 TCTTTGGTTCAGCCTTCCCTTGGCAGCTTGCTTGGCTGTTGGGGCAGCCCTTGGTCCTAC
 GGACTTGGTGGCCTTTGCCTCTCTTTTCGGAGCGTTTTAGCTTTCCTAAGCGCGTGTCCAA
 TATTCTTAAGGGCGAAGGACTCTTGAATGATGCTTCTGGTTTGGTGGCTTTTCAGGTAGC
 TTTGACAGCTTGGACAACCTGGAGCTTTTTCTCTGGGGCAAGCTAGCAGTTCGCTCATCTT
 TTCAATCCTAGGCGGTTTTTTAATTGGATTTTTAACAGCCATGACCAACCGCTTCCTCCA
 TACCTTCTTGCTAAGTGTGCGCGCAACGGATATTGCCAGTGAACTTTTATTAGAATTGGA
 GTTTGCCTCTAGTGACCTTCTTTCTGGCAGAAGAAGTCCATGTTTCAGGGATTATTGCCG
 TCGTAGTTGATCGAATTTTAAAGGCAAGTCGCTTCAAGAAAATCACGCTCCTCGAAGCCC
 AAGTGGATACGGTGACCGAGACGGTCTGGCATAAGTGACCTTTATGCTCAACGGTTCTG
 TCTTTGTGATTTTAGGGATGGAGTTGGAATGATAGCAGAACCTATCTTGACCAATCCAA
 TCTATAATCCTCTACTTTTATTGCTATCTCTCATCGCCCTTACCTTTGTCTCTTTGTCA
 TTCGTTTTATTATGATCTATGGCTATTATGCCTATAGAACCCGACGCTAAAGAAAAAGC
 TAAATAAGTATATGAAGGACATGTTTCTCTTGACCTTTTCAGGTGTTAAGGGAACGGTGT
 CGATTGCTACGATTCTCTTGATACCAAGTAATCTAGAACAGGAGTATCCTCTCTTGCTTT
 TCCTTGTTGCAGGTGTGACGCTTGTCAGCTTTTAAACAGGTCTCTTGGTCTTGCCTCATC
 TTTCTGATGAAGAGGAAGAAAGCAAGGATTATCTCATGCATATCGCCATTTTGAATGAAG
 TAACGCTAGAGTTGGAAAAAGAGTTGGAAGACACCAGAAATAAACTTCCCCTCTATGCGG
 CTATTGACAATTCGATCATGGACGTATTGAAAATCTCATTTTAAAGCCAAGAAAACCAGGA
 TGATCAAGAAGACTGGGCTGCTTTGAAAATCGAATTCTTAGTATTGAAAGTGATGGTTTG
 GAACAGGCCTATGAAGAGGGGAACATTAGCAATCGTGCTTACCGAGTTTACCAACGTTAT

CTGAAAAATATAGAACAAGGAATCAATCGTAAACTTGCCTCAAGACTGACCTATTATTTT
CTTGTTTCCTTGAGGATTTTACGTTTTCTTCTTCATGAAGTTTTTACTCTTGGAAGACC
TTCCGTAGCTGGAAGGACAAGGAGCAAAGCCGTCTCCGTGCTCTTGATTATGACCAAATT
GCAGAGCTCTATCTTGCCAATACAGAGATGATTATTGAAAGTTTGGAACCTGAAGGGA
GTCTACAGACGCTCTTTGATTAGTTTTATGCAGGAGTCTCGTCTTCGAGAAACAGCTATT
ATCAGCAGTGGTGCCTTTGTGCAACGGGTATCAATCGTGTCAAACCCAACAATATCGAT
GAAATGCTGAGAGGCTATTATCTGGAGCGCAAGTTGATTTTTCGAATACGAAGAAAAACGA
TTGATTACGACTAAGTATGCCAAGAAATTACGACAAAATGTAAATAACTTAGAGAACTAT
TCCTTGAAGGAAGCTGCCAATACCCTGCCGTATGATATGGTGGAATTGGTAAGAAGAAAT
TAGTTAATACTCTTCGAAAATCTCTTCAAACCACGTCAGCGTCGCCTTGGATTATATATG
TGACTGACTTCGTCAGTTTCATCTACAACCTCAAAGCAGGGCTTTGAGCAACCTGCGGCT
AGCTTCCTAGTTTGCTCTTTGATTTTCATTGAGTATAAGATTGTAAGTGAAGGAGTGTGA
CATGAAAAAATGGGGAAAGAGCCTGAACTAGTCTGTCTACTTTTACCCAATCACACTTC
CATTTGGTACAGCTGGATCAACTGTGAGAAGGGATCGAATTTGCCATCATGTTTCAGCTGA
GAGAATCATACCCTGGCTGACATATTTTTTTCATCATTTTACGTGGTTTGAGGTTAGCAAC
GATTTGAACTTTCTTGCCGACCAATTCTTGTTTCATTTGGATAGTATTTTGCAATTCCTGA
AAGAATCTGACGATCTTCTCCATCACCAGCATCCAAGCGGAATTGAAGCAACTTATCTGA
ACCTTCTACTTTAGACACTTCTTTGACTTCTGCGACACGGATTTCAACCTTGTCAAAGTC
TTCAAACCTTGATTTTCATCCTTGTTTTAGTTTGAGCTCAACTTCGTCCGGATTCCATTCTTT
TTCGACTGCTGGTTTTATTGCCTTCCATTTGTTCCTTGATATAGGCGATTTCTTCTTCCAT
ATTTAGACGTGGAAAGATAGGTGTTCTTTGGCAACTACAGTCACATCTGCTGGGAAGTC
AGCCAAACTCAAGTTTTCAAGACTAGAACTTCTTCCAAACCAAGTTGAGTCAAACTGC
ACGACTAGTTTCCATCATAAATGGTTCAATCAAGTGAGCAACTACACGAATGCTGGCTGC
CAAGTGGCTCATGACACTTGCCAATTGGTCACGAAGAGCTTCATCCTTGTTCCAAGACCCA
TGGTGCAGTCTCATCGATGTATTTATTGGTACGAGAGATCAGAGTCCAGACTGCTTCAAG
CGCACGTGGATAGTCAACTGCTTCCATGTGTGTATGGAAGTCTGCGATTGATTTTTCTGC
AACCTCAGCAAGAACATGATCAAATTCAGTCACACCTTCTACATAGGCAGGGATTTGTCC
ATCAAAGTACTTATTAATCATGGAAACCGTACGGTTAAGGAGGTTCCCAAGGTCATTAGC
CAATTCATAGTTGATACGACCGACATAGTCTTCAGGAGTAAAGGTTCCGTCTGAACCAAC
TGGAAGGTTACGCATGAGGTAGTAACGAAGTGGATCTAGTCCATAACGCTCTACCAACAT
TTCAGGGTAAACGACATTCCCTTTTGACTTAGACATTTTTTCCGTCTTTCATGACAAACCA
ACCATGGGCAATCAAACGATCAGGTAATTTAACATCCAACATCATAAGAAGGATTGGCCA
GTAGATAGAGTGGAAGCGAAGGATGTCTTTTCCCTACCATATGGAAGACTGTTCCATTCCA
GAACCTGTCAAAGTTACCATGTTTCGTCTTGAGCGTAGCCAAAAGCTGTCGCATAGTTAAG
AAGGGCATCAATCCAAACGTAGACAACGTGTTTTGGATTTGATGGGACAGGCACTCCCCA
TGTAAGGTTGTACGAGATACCGCCAAATCTTCCAAACCTGGCTCGATGAAGTTGCGTAG
CATTTTCATTAAGACGACCATCTGGCGTGATAAATTCAGGATGAGCTTTGAAAAATTTCGAC
CAAACGGTCTTGGTATTTGCTAAGGCGAAGGAAGTATGATTCTTCAGAAACCCATTCAAC
CTCATGACCTGATGGAGCAATACCACCAGTCACATTTCCAGCTTCATCACGGAAAACTTC
TGCCAGCTGGCTTTCTGTAAAGAATTCTTTCGTCTGATACTGAATACCAACCAGAGTATTC
ACCCAAGTAGATATCATCTTGAGCAAGTAAGCGTTCAAAGACCTGTGCGACAACCTTTTC
ATGGTAGTCATCGGTTGTACGGATAAAATTTATCGTATGAGATATCTAGTAATTGCCAGAG
TTCTTTAACTCCAACCGCCATTCCATCAACATAGGCTTGAGGTGTAATACCAGATTCGAA

TTCCGCTTTCTGCTGGATTTTCTGACCATGTTTCATCAAGACCTGTCAGATAAAATACATC
 GTAGCCCATCAGGCGTTTGTAACTGCTAGGACATCACATGCGATAGTTGTGTAGGCAGA
 ACCGATATGAAGTTTCCCAGATGGATAGTAAATCGGCGTTGTAATATAAAAATTTTTTTC
 AGACATAATTTTTCCTTTCCAGGCAAATGAAACCTGTTTTTCTAACACTTCATTATATCA
 CATTTTAAATGAATTTTCGATAGGGAAATCCATACCAAAACAAGATAGACGAGTGTCCATC
 TTGTTGATCTCATTCATAACGAAGGGCTTCAATTGGATCAAGTTTCGATGCCTTGTTGGC
 TGGCAAGACTCC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 197 | 670 | F | 158 aa |
| 2 | 612 | 1304 | F | 231 aa |

> 3864184-1 ORF translation from 197-670, direction F
 VIFISTLSLGLAHLWFSPLAACLAVGAALGPTDLVAFASLSEFSPKRVSNILKGE
 GLLNDASGLVAFQVALTAWTTGAFSLGQASSSLIFSILGGFLIGFLTAMTNRFLHTFLLS
 VRATDIASELLEFEFASSDLLSGRRSPCFRDYCRRS*

Description:

unknown

> 3864184-2 ORF translation from 612-1304, direction F
 VTFFLAEEVHVSGIIAVVVDRIKASRFKKITLLEAQVDTVETVWHTVTFMLNGSVFVI
 LGMELEMLIAEPILTNPINPLLLLLSLIALTFVLVIRFIMIYGYAYRTRRLKKLKNKY
 MKDMFLLTFSGVKGTVSIATILLIPSNLEQEYPLLLFLVAGVTLVSFLTGLLVLPPLSDE
 EEESKDYLMHIAILNEVTLELEKELEDTRNKLPLYAAIDNSIMDVLIKISF*

Description:

unknown

Assembly ID: 3864194

Assembly Length: 1941bp

> 3864194 Strep Assembly -- Assembly id#3864194

AATTAGTATTCTCAACCTTTTTATCTTGATAGTTCAAGATGGCATTCGTTGAATTGGTAA
 CATAGTAACATCCACTCCCTTCAGTTTAGCTGCCTCTTGAACCCAGGATCTTGCGGTT
 TTGGCGGTTCAACAGGAATTCTTTTCTTTTCCAGAAACCGTAAAAGCTGATTGTTTCTG
 AGTAAAAGACCCATCTTTACTTTTTTTAGGAGAGAAAAAGACGCTAATATTTTTCTGAGA
 TTTAGTCATATCTTTATTGACTTGACGAGATAGGGAATCACCCAAAGCCATAATCACAAC
 AACTGATGAAACACCGATAATAATCCCAATCATAGTAAGCAAAGAACGCATCTTGTGAGC
 CATGATAGATGAAAAGGCAAATTCAGATTCTGCATCTTAGTTTTCTCCTTTCTTAACCT
 GAGCACTGTCAGACGAAATGACCCCATCCCGAATGACAATCTGACGTTTGGCATAGGCAG

CAATCTCAGGCTTCATGCGTTACCATGATAATGGTTTTTCCTTCTTTATTCAAATCAACC
 AATAATTGCATAATTTGGTTACCTGTTTTGGTATCCAAGGCTCCTGTCGGTTTCATCCGCT
 AGGATAATAGAAGGATTGTTTACCAAGGCACGCGCAATGGCTACACGTTGCTTTTGACCA
 CCAGATAATTCTGAAGGTAAATGGTGACTACGTTCTATCAATTCAACCTTGTCTAAATAT
 TCCTCAGCCAACTTGCGACGTTTTGAAGACGAACTCCTGCGTAAATCAAGGGCAATTCT
 ACATTTTGCAGAGCATTGAGCTTCGATAGAAGAAAGAACTGCTGAAAAGACAAAACCGATT
 TGTGTTGGTTACGGACCTTAGCTAGTTGTTTTTTCACCAAGCCCAGCCACTTCTTGACCTTCA
 AGATAATATTCTCCACTGGTTGGTGTATCCAACATGCCAATCGTATTCATCAGAGTGGAC
 TTACCAGACCCAGATGGTCCCATGATGGCTACAAATTCACCCTCATTCACTTCTAGATTG
 ATATTTTTTGAGAACCTGCAGTTCTTGGTCACCATTACGGTAACTTCTGAAGATATTTTTT
 AGACTAATTAGTTGCTTCATCAGCCTTCACCTCTTTTCCTTCTTCCAAGGAAGATGTTGG
 ATTACTGATGACCTTAGCACCGTTTCGTTAAACCAGAAGTGATTTCTTGATTTTCTGCGTC
 AGCATTTCCCAATGAAACCTCAACTTTTTTAGCCTTTTGTGTTTCATCCACAATCCAGAC
 ATAATTTTTTACTATCATCCATTACTAGACTGCTAACAGGAACAAGAATAGCCTTAGTTTT
 GCTTTTAAACCTCAATGTTGACAGAAAAACCTTGTTTCAAATCACCAACCTCGCCTGTCAC
 ATCAATAGTATAAGGGTATTTAGAACCTGTATTATTTCCCGGCTGCTGGACTAGCTGCTTC
 ACCATTGTTTTTTAGGATAGTCAGAAATATAGGCTTAATTTCCCAGTCCATTTTTTTATCAG
 GATACACTTTAGAAGTAAAGCTTACTTCTTGACCTACAGAAAGTTGGCTAGATTGTACT
 CAGACAATTCTCCCTTGACTTGTAATTTTCATTGCTGACAATATGAACCATAACTTGAC
 TCGCCCCCTGTTGGAGATTTAGAAACATTGCTATTGACTTCGACTACAGTTCCCTCTAGGG
 TACTGAGAACAGTTGTTGCATCCAATTGACTTTGAGCCTTGCTTAATTGCGCTGCAGCAT
 CTGCACGCGCATCACGGGCATCACCCAATTGAGCATCAATAGAAGCAACAGAATTTCCAG
 CCACTGGAGTTGGGCTTTGCACCGTTGCATCTTCTCCTCCTACTGGCGCTGGTAACTGTG
 GAGCCTGAGCTGAAGCGGCTTCATTTTCGTGCTTGATTGAGTTCATTGATATGACGATCTG
 CCTTAGCTACTGCTCGACTAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1084 | 1380 | R | 99 aa |

> 3864194-3 ORF translation from 1084-1380, direction R
 VTGEVGLKQGFVSVNIEVKSKTKAILVPVSSLVMDSDKNYVWIVDEQQKAKKVEVSLGNA
 DAENQEITSGLTNGAKVISNPTSSLEEGKEVKADEATN*

Description:

unknown

Assembly ID: 3864338

Assembly Length: 1335bp

> 3864338 Strep Assembly -- Assembly id#3864338

ATCGAATTCCTATTTTAACTTTCTTTTCTAAAACAGTCTATATTTTATTTCAAACCTG

TATTATATTTTTGAAAAATAAAGTCCTTTTTTCTTTTTTTCAGAAAAAGGGTATAATA
AAAGAAAAATAAGCAGTAACACTCAATGGAAATCGAAAAAGCAAAGCTAGGAAGCTAGCCGC
AGATTGCTCAAAACACTGTTTTGAGGTTGCAGATAGAGCTGACGTGGTTTGAAGAGATTT
TCGAAGAGTATAAAAAGGTGCTAGGCATGTTGATTTTTTCCTTTGTTAAATGATTTGTCAA
GAAAAATCATCCATATTGGACATGGATGCCTTTTTTGCTGCAGTGGAAATCAGGGATAAT
CCTAAACTCAGAGGAAAACCTGTCATTATTGGAAGCGACCCCTCGGCAAACAGGTGGACGG
GGAGTCGTTTCTACCTGTAGTTATGAGGCAAGAGCTTTTGGTGTCCATTCTGCCATGAGT
TCCAAGGAAGCTTATGAACGTTGTCCCCAGGCTGTCTTTATCTCAGGGAATTCGATGAGA
AATACAAGTCTGTGGGACTCCAGATTCGAGCTATTTTTAAGCGCTATACAGATTTGATTG
AACCCATGAGCATTGACGAAGCCTATTTGGATGTGACAGAAAATAAACTCGGTATCAAGT
CAGCGGTCAAATGCTCGCCTCATTCAAAAAGATATCTGGCAAGAACTCCATCTAACTG
CTTCCGCAGGCGTTTCTTACAACAAATTCTTAGCTAAAATGGCGAGTGATTATCAAAAAC
CACATGGTTTTGACAGTGATTCTACCTGAACAGGCTGAGGATTTTCTCAAACAAATGGATA
TTTCCAAATTTTCATGGAGTAGGAAAAAAGACAGTAGAACGTCTTCATCAAATGGGCGTTT
TTACTGGTGTGATTACTTGAAGTTCCTGAGGTAACCCTAATAGACCGTTTTGGTAGAC
TAGGCTATGATCTGTATCGAAAGGCTCGTGGCATTACAACTCTCCAGTCAAATCCAATC
ACATCCGTAAATCAATCGGCAAGGAGAAAACCTACGGGAAGATTCTCCGTGCTGAGGAAG
ATATCAAAAAGAGAGCTGACTCTTCTATCAGAAAAAGTCGCTCTCAATCTACATCAACA
AGAAAAAGCTGGAAAAATTGTCATTTTGAAAATCCGCTACGAGGACTTTTCAACTCTTAC
CAAACGAAAAAGTATTGCTCAAAAACACAAGATGCTAGTCAGATAAGCCAAATAGCCCT
GCAACTCTATGAAGAATTAAGTGAGAAAGAAAGAGGTGTCCGCCTATTGGGGATTACCAT
GACTGGATTTTAAAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 552 | 1100 | F | 183 aa |

> 3864338-2 ORF translation from 552-1100, direction F
VGLQIRAIKRYTDLIEPMSIDEAYLDVTENKLGKISAVKLIARLIQKDIWQELHLTASAG
VSYNKFLLAKMASDYQKPHGLTVILPEQAEDFLKQMDISKFHGVGKKTVERLHQMGVFTGA
DLLEVPEVTLIDRFGLGYDLYRKARGIHNSPVKSNHIRKSIGKEKTYGKILRAEEDIKK
ES*

Description:

ECODINJ NCBI - Escherichia coli (sub_strain W3110, strain K-12)
DinP, DNA damage inducible protein

Assembly ID: 3864360

Assembly Length: 1796bp

> 3864360 Strep Assembly -- Assembly id#3864360

TCCAAGCTAGCTATTTCTGTTGAAGGGGCTTCGGTTGGCAGAACCTGGTGAATTTACCCAA

ACGTGCTTTTTTAAACGGTCGCGTAGACTTGACACAGGCAGAGGCTGTGATGGATATCAT
 CCGTGCCAAGACTGACAAGGCCATGAACATTGCGGTCAAACAATTAGACGGCTCCCTTTC
 TGACCTCATTAACAATACCCGTCAAGAAATCCTCAATACACTTGCCCAAGTTGAGGTCAA
 TATCGACTATCCTGAATATGATGATGTTGAGGAAGCTACTACTGCCGTTGTCCGTGAGAA
 GACTATGGAGTTTGAGCAATTGCTAACCAAGCTCCTTAGGACAGCACGTCGTGGTAAAAT
 CCTTCGTGAAGGAATTTCAACGGCTATCATTGGACGTCCCAACGTTGGGAAATCAAGCCT
 TCTCAACAACCTCTTGCGTGAGGACAAGGCTATCGTAACCGATATCGCTGGGACAACACG
 AGATGTCATCGAAGAGTACGTCAACATCAATGGTGTTCCTCTAAAATTGATTGACACAGC
 TGGTATTTCGTGAAACGGATGATATCGTTGAACAAATCGGTGTTGAGCGTTCGAAAAAAGC
 CCTCAAGGAAGCCGACTTGGTTCTACTAGTGCTAAATGCCAGTGAACCACTGACTGCGCA
 AGACAGACAACTTCTTGAAATTAGCCAAGATACCAATCGCATTATTCTACTTAATAAAAC
 CGACCTGCCAGAAACGATTGAAACTTCGAAACTACCTGAAGACGTTATCCGTATTTTCAGT
 CCTTAAAAACCAAACATCGACAAGATTGAAGAGCGAATCAACAACCTCTTCTTTGAAAA
 TGCTGGCTTGGTCGAGCAAGATGCTACTTACTTGTCAAACGCCCGTCACATTTCCCTGAT
 TGAAAAAGCAGTTGAAAGCCTACAAGCCGTTAATCAAGGTCTTGAGCTGGGGATGCCAGT
 TGATTTGCTTCAAGTTGACTTGACTCGTACTTGGGAAATCCTCGGAGAAATCACTGGGGA
 TGCTGCTCCAGATGAACTCATCACCCAACTCTTTAGCCAATTCTGTTTAGGAAAATAAGA
 AAAATCCATGATCCTTCATTTCGGTCATGGATTTTATTGTCTTTATTAGTAATCTGGTCTT
 AAGACCCCTGTTACAGTTGCCTTAGTTGCTTCGTAGTCGCCATCTACGACAACCTTGATA
 ATGCGTTTGACATCTTCTTCTGGTGCTGGAACAAGAGGTAGACGAGTGGGTCCAGCTTCA
 AATCCCATATAGTTAAGAATTGCCTTAACTGGAGCAGGACTTGGATAAGAGAAGAGAGCA
 TTAACCTTAGGAATGAATTTACGCTGAATTGCTGCGGCTTCTTCATATCGCTTTCTGCA
 ATGGCAGTAAACATCTCGTGCAATTTTCATCCCCATTTGTATGAGAGGCAACAGAAATAACC
 CCATCCGCCCCAAGGTTTCATGGCATGGAAAGCATCTCCATCCTCACCTGTATAAATCAAG
 AACTCTTCAGGCTTGTGCTCAATCAAGTAAGCCATATTAGCCAAGCTAGTACATTCTTTG
 ACACCGATAATATTTGGATGGTCAGCCAAGCGAAGCATGGTTTCTGGAGTCAATTGACA
 ACTACACGCCCTGGAATGTTATAGATAATAATTGGTAGGTCAGAAGCATCTGCAATAGCC
 TTAAAGTGCTGATACATCCCTTCTTGAGAAGGTTTGTGTAGTAAGGAACAATAGCAAGC
 CCAGCTGCGAAACCACCAAATTCGCTACTTCTTTGACAAACTCAATAGAGTCACG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 47 | 1078 | F | 344 aa |

> 3864360-1 ORF translation from 47-1078, direction F
 VNLPKRAFLNGRVDLTQAEAVMDIIRAKTDKAMNIAVKQLDGSLSDLINNTRQEILNTLA
 QVEVNIDYPEYDDVEEATTAVVREKTMEFEQLLTKLLRTARRGKILREGISTAIIGRPNV
 GKSSLLNNLLREDKAIVTDIAGTTRDVIEEYVNINGVPLKLIDTAGIRETDDIVEQIGVE
 RSKKALKEADLVLLVLNASEPLTAQDRQLEISQDTNRIILLNKTDLPETIETSKLPEDV
 IRISVLKNQNIDKIEERINNLFENAGLVEQDATYLSNARHISLIEKAVESLQAVNQGLE
 LGMPVDLLQVDLRTWEILGEITGDAAPDELITQLFSQFCLGK*

Description:

THIOPHENE AND FURAN OXIDATION PROTEIN THDF. - ESCHERICHIA COLI.

Assembly ID: 3864388

Assembly Length: 2337bp

> 3864388 Strep Assembly -- Assembly id#3864388

```
CTTCGTACAGGTGGTTTCCTATGCAAGGGTGGGAAGCCAATCGTCAGAACAACAAGCATCTT
CATCAAGCCAGAACTGGAGCAATTACAAAAAGAAATTGCTGAAGAAGAAGCAAGCTTGGG
TTCAGAAGAAGTGGCTTTGAAGACCTTGCAAGATGAGATGGCCAGATTGACCGAGTCATT
AGAAGCTATTAAATCTCAAGGAGAGCAGGCACGTATTCAGGAGCAAGGCTTGTCCCTCGC
TTATCAGCAAACTAGTCAGCAAGTTGAAGAACTGGAACTCTTTGGAACTCCAAGAAGA
GGAAATAGATCGTCTTTCCGAGGGAGATTGGCAAGCGGATAAGGAAAAATGCCAAGAGCG
TCTTGCTGCAATCGCCAGTGACAAGCAAAATCTGGAAGCTGAGATTGAAGAGATTAAGTC
TAATAAAAATGCCATCCAAGAACGCTATCAAACTTGCAAGGAAGAGCTAGCGCAAGCTCG
TTTGCTTAAGACAGAACTGCAAGGGCAAAAACGTTATGAAATTGCTGATATTGAACGCTT
AGGCAAGGAATTGGACAATCTTGATTTTGAACAAGAGGAAATCCAGCGCCTTCTTCAAGA
AAAGGTTGACAACTTGAGAAGGTTGATACAGAATTGCTCAGTCAACAGGCGGAAGAATC
CAAACTCAGAAAACGAACCTCCAACAAGGTTTGATTTCGCAAACAGTTTGAGTTGGATGA
TATAGAAGGTCAGCTGGATGATATTGCTAGTCATTTGGATCAGGCTCGCCAGCAGAATGA
GGAGTGGATTTCGCAAGCAAAACACGTGCTGAAGCTAAGAAAGAAAAGGTCAGCGAGCGCTT
TGCCGCCATCTACAAAGTCAATTAACAGACCAGTACCAGATTAGCCATACTGAAGCTCTA
GAAAAAGCGCATGAATTGGAAAACCTCAATCTGGCAGAGCAAGAAGTTAAGGATTTAGAG
AAGGCTATTTCGCTCACTGGGTCCTGTCAATATAGAAGCTATTGACCGGTACGAAGAAGTT
CACAACCGTCTGGACTTTCTAAATAGTCAGCGAGATGATATTTTGTCTAGCGAAAAATCTG
CTCCTTGAAACCATTACAAAGATGAATGATGAGGTTAAGGAACGCTTTAAATCAACCTTT
GAAGCTATTTCGTGAGTCCTTTAAAGTGACCTTCAAGCAGATGTTTGGCGGAGGTCAGGCA
GACTTGATATTGACTGAGGGCGACCTTTTACAGCTGGTGTGGAGATTTCTGTTCAACCTC
CAGGTAAGAAAATCCAGTCGCTTAACCTCATGAGTGGTGGTGAAAAAGCCCTATCGGCTC
TTGCCTTGCTTTTCTCCATTATTTCGTGTCAAGACCATTCTTTTGTCTCTTGATGAGG
TGGAAGCTGCGTTGGATGAAGCCAATGTTAAACGTTTGGGGATTACCTCAACCGCTTTG
ACAAGGACAGCCAGTTTATCGTCGTAACCCACCGTAAGGGAACCATGGCAGCGGCCGATT
CCATCTATGGAGTGACCATGCAAGAATCGGGTGTTCAAAGATTGTTTCAGTTAAGTTAA
AAGATTTAGAAAGTATTGAAGGATGACAATTAACTAGTAGCAACGGATATGGACGGAAC
CTTCCTAGATGAGAATGGGCGCTTTGATATGGACCGCCTCAAGTCTCTCTTGTTTCTTA
CAAGGAAAAAGGGATTTACTTTGCGGTGGCTTCGGGTCGGGGATTTCTGTCTCTGGAAAT
CGAATTATTTGCTGGTGTTCGTGATGACATTATTTTCATCGCGGAAAATGGCAGTTTGGT
AGAGTATCAAGGTCAGGACTTGTATGAAGCGACTATGTCTCGTGACTTTTATCTGGCAAC
TTTTGAAAAGCTGAAAACGTCACCTTATATAGATATCAATAAACTGCTCTTGACGGGTAA
GAAGGGTTCATATGTTCTAGATACGGTTGATGAGACCTATTTGAAAGTGAGTCAGCATTA
TAATGAAAATATCCAAAAAGTAGCGAGTTTGGAAGATATCACAGATGACATTTTCAAAAT
TACAACCAACTTCACAGAAGAAACGCTAGAAGCTGGTGAAGCTTGGGTCAATGATAATGT
CCCTGGTGTCAAGGCTATGACAACTGGCTTTGAATCTATTGATATTGTTCTGGACTATGT
```

CGATAAGGGTGTAGCTATTGTTGAATTAGCTAAAAAACTTGGCATCACAATGGATCAGGT
 CATGGCTTTTGGAGACAATCTTAATGACTTACATATGATGCAGGTTGTGGGACATCCTGT
 AGCTCCTGAAAATGCACGACCAGAGATTTTAGAATTAGCATAAGACTGTGATTGGTC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1239 | 1586 | F | 116 aa |

> 3864388-3 ORF translation from 1239-1586, direction F
 VEISVQPPGKKIQSLNMSGGEKALSALALLFSIIRVKTIPIFVILDEVEAALDEANVKRF
 GDYLNRFDKDSQFIVVTHRKGTMMAADSIYGVTMQESGVSKIVSVKLDLESIEG*

Description:

P115 protein - *Mycoplasma hyorhinis* (SGC3) (similarity to
 SMC1_YEAST, chromosome segregation protein)

Assembly ID: 3864406

Assembly Length: 2162bp

> 3864406 Strep Assembly -- Assembly id#3864406

CTAAAAGTGAAGCCCGATAGCGTCTCTCTCTGCAAGGATTCATAACCAATAACAGGAG
 ATTGACGAACAATAATCGGTTGAATGACCCCATTTTCTTTGATAGACTGTGCTAGTTCAT
 CTAGCTTTTCTCTATCAAATCTTTTTCGGGTTGATAGGGATTTTTTTGTATATCTGTGA
 TAGAAATCATTTCAAATTTTCCATGATTCTACACTAACACATCTTTTCTCTTATGTAAA
 GCTTTCTTTACATAGATGTCAATTAAGATTCTAAATCACCTGAACCTCTTGTTAAGTTTGA
 TAGAGGTAGTTTCTTCTTTCCCGTTACGATAGTAGGTTATCTTAATGGTGTCTCCGATAG
 AATGGTTGTAAAGAGCACTTTGTAAGTCTGTTGATGAAGCAATCTCTTTGTCTACTT
 TTGTAATTACATCGTATTTTCAAGGTGACCATTGGCAGGCATATTACTTTGTACCGAAC
 GAACAATTACACCAGATGTAACATTACTTGGAATATTGAGTCTTCTGATGTCGCTTGTA
 TCACATTAGATAAATTAACCATCTGGATTCCCAAAGCTGGACGCGTCACTTTTCCGTTTT
 TTTCTAACTGTTCAATAATATTGATAGCATCATTTGCAGGAATTGCGAAACCAAGACCTT
 CTACAGATGTTCCCTCCATTTGTAGCAATTTTACTTGAGGTAATTCCGATAACCTGCCCTT
 GAATATTGATCAGTGGGCCGCCAGAGTTACCTGGGTTAATAGCAGTATCAGTTTGGATGG
 CTTTTGTAGAAATAGCTTGTCCATCTTCCGATTTTAAGGATACATTTCTATTGAGACTGG
 ATACGATACCTTGAGTGACAGTATTTGCATATTCAGAACCTAACGGGCTACCGATGGCAA
 TAGCAGTTTCTCCTACAGTTAACTTACTAGAATCACCAAACCTCAGCTACTGTTGTCACTT
 TTTCTGAAGAGATTTTCGACGACAGCAATATCAGAGAAAGTGTGAGCTCCGACAATTTCTC
 CAGGTACTTTAGTCCCATCTGACAATCGAATATCTACTTTGCTGGCGCCATTTATAACGT
 GATTGTTGGTGACGATGTAAGCTTCTTTATCATTTCTTTTATAAATAACTCCAGATCCTT
 CACTAGAGATTGCTGAGAATCTGTGTCAGTATCATCATTTGCCAAATACGCTATTTGTGTC
 TGTTTGGCCGAATAAGTAATAACAGAAACAACAGCATCTTTTACTTTGTTAACGGCCTGTG
 TTGTTGAATTTTCCGTTCCCTTATAGGCAGTTTGTGTAATAGTACTATTGTTGTTAGAGTT

GTTTACACTACTTTTTTTGAGTTAGTTGAGTTATTGAAAACTACCCAAGGCTCCACTAAA
 AAAGCTAATGACGATAACGACTAATAATTGAAACCATTTTTTTGTAAAATGTTTTTAGATG
 TTTCATATTTGCCTCCATATGTTTGAATTACTGAAAGTATAAACTGACTAGCTTAATTAT
 AACTTAAACACAAAAGTTTTTACACAACTGTGGATAACTCTTTTGAACTGTGATTTTCT
 TAATTGAAATCTATTTTTTATTTTGTGAATAAGATGTGAAAAATAGAGAATATGTTAGA
 ATAGAGTCATGAAAATTAAAGTTGTAACAGTTGGGAACTGAAAGAAAAGTATTTAAAAG
 ATGGTATCGCAGAGTATTCAAACGAATTTCTAGATTTGCTAAGTTTGAAATGATTGAGT
 TATCAGATGAAAAAACACCAGATAAGGCCAGTGAATCAGAAAATCAAAGATTTTAGAAA
 TAGAAGGTCAGAGAATTTTATCAAAAATTGCTGACCGTGATTTTCGTTATTGTGTTAGCCA
 TTGAAGGGAAAACTTTCTTCTCAGAAGAATTTAGTAAGCAGTGAGAAGAACTTCTATAA
 GGAAGGATGTCTACTCTTACTTTTTATTATTGGGGGAAGTTTAGGATTGTCATCATCTGTA
 AAAAATAGAGCCAATCTTCTGTGTCAGTTTGGTTCGCCTAACCTTGCCTCATCAGTTAATG
 AGACTAGTTCTTGTGTAACAAATCTATCGCGCTTTTACGATTCAGCAGGGATTCCCCTAC
 CATAAATAGAGAATTGACTTTTAATTGAATTTTTTGGTAGAATAATTGTGTTAGGTCTCAT
 AG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 263 | 958 | R | 232 aa |

> 3864406-1 ORF translation from 263-958, direction R
 VTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGVSSSLNRNVSLKSEDGQAISTKA
 IQTDTAINPGNSGGPLINIQQQVIGITSSKIATNGGTSVEGLGFAIPANDAINIIEQLEK
 NGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIVRSVQSNMPANGHLEKYDVITK
 VDDKEIASSTDLQSAALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES*

Description:

Bacillus subtilis (strain 168,) DNA. Homologous to E. coli
 serine protease HtrA (BLAST)

Assembly ID: 3864452

Assembly Length: 1766bp

> 3864452 Strep Assembly -- Assembly id#3864452

ATCGAATTTTCCAAAATGGGGAGCTAGAGCAGTGGAGTGATTATGTGGCAGACGATTTGA
 TTCAGCATAATCATGAGATTGGACAAGGAAGTGCTGCTTATAAAAACATATGTGGCTGAAT
 ATATTGTCACTTTTGACTTCGTTTTTCCAACCTCTTAGGACAAGGAACTATGTGGTTAGCT
 ATGGTCAGACTCAGATTGATGGCGTTGCTTATGCCAAGTACGATATCTTCCGTTTAAAGA
 ACGGGAAAATTGTGGAGCATTGGGATAATAAGGAAGTCATGCCTAAGGTAGAAGACTTGA
 CCAATCGAGGGAAGTTTTAAATTGAGGACAAAGAATGATTGAATACAAAATGTAGCACT
 GCGCTACACAGAAAAGGATGTCTTGAGAGATGTCAACTTACAGATTGAGGATGGGGAATT
 TATGGTTTTAGTAGGGCCTTCTGGGTCAGGTAAGACGACCATGCTCAAGATGATTAACCG

TCTTTTGGAAACCAACTGATGGAAATATTTATATGGATGGGAAGCGCATCAAAGACTATGA
 TGAGCGTGAACTTCGTCTTTCTACTGGTTATGTTTTACAGGCTATTGCTCTTTTTCCAAA
 TCTAACAGTTGCGGAAAATATTGCTCTCATTCCTGAAATGAAGGGGTGGAGCAAGGAAGA
 AATTACGAAGAAAACAGAAGAGCTTTTGGCTAAGGTTGGTTTACCAGTAGCCGAGTATGG
 GCATCGCTTACCTAGTGAATTATCTGGTGGAGAACAGCAACGGGTCGGTATTGTCCGAGC
 TATGATTGGTCAGCCCAAGATTTTCCTCATGGATGAACCCTTTTCGGCCTTGGATGCTAT
 TTCGAGAAAACAGTTGCAGGTTCTGACAAAAGAATTGCATAAAGAGTTTGGGATGACAAC
 GATTTTTTGTAACCCATGATACGGATGAAGCCTTGAAGTTGGCGGACCGTATTGCTGTCTT
 GCAGGATGGAGAAATTCGCCAGGTAGCGAATCCCGAGACAATTTTAAAAGTGCCTGCAAC
 AGACTTTGTAGCAGACTTGTGGAGGTAGTGTTCATGACTAATTTAATTGCAACTTTTC
 AGGATCGTTTTAGTGATTGGTTGACAGCTACAATGACATTGGTCGGTTCCTTGAGCAAGA
 GATAGATTAGCCAGACAGTCATGCCCCAAATCCCTCCAGGTAAGAGCATAGACCGTTGCA
 CATTAAGTACGATTAAAAAAGTGATAATGGCAAGAAAACCTTGCTACTGCTTGTAAATAAAA
 AGGTTGTTAGTGTTCATATTAGTTCATCAATACCAAGGCGACAGAAGTTCCTGCCCTAAA
 GCGAGGGTAATGAGCAGGGATTCAAACATCTTACTCATACCAGAGTTTATGTGGTTGGTC
 ATAATATCACGGACCGCATTTGGTCAAGGCAATACCTGGTACAAACGGCATGACCGCACCA
 GCTATAATCAAATCTGCCGTTGAAGGAAAACCTGTGTAGCGAGCCCCAAACTGGGCAATT
 ATCCCAAAGACAAAAGCTCCAGCAAAGGCTGTCAAAAGGGAATTCGGATAAATTTTCC
 ACATAGAGGGAAAAGGCAAAACCAAATAAGGTCGCCACTCCTGCCCAAGTGCCTCGTAG
 ATATTTCCGCTAAACATAACTGAAAAGAAAGGAGCACTAAAGGTGCGAGCCAGAGTTACC
 TGCAACTTAGTATAGGGAAGGGGTTGAGCTTGCAAGGCCGTCAATTGCTTAAAGGCTGTT
 TCTAAGTCAATCTGCCCCCAACTGG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1079 | 1201 | R | 41 aa |

> 3864452-2 ORF translation from 1079-1201, direction R
 VQRSM LLPGGILGMTVWLIYLLLKEPTNVIVAVNQSLKRS*

Description:

unknown

Assembly ID: 3864458

Assembly Length: 1705bp

> 3864458 Strep Assembly -- Assembly id#3864458

CTCTGACGGAGGCTGGTTATGTGGGTGAGGATGTGGAAAATATACTCCTCAAACCTCTTGC
 AGGTTGCTGACTTTAACATCGAACGTGCAGAGCGTGGCATTATCTATGTGGATGAAATTG
 ACAAGATTGCCAAGAAGAGTGAGAATGTGTCTATCACACGTGATGTTTCTGGTGAAGGGG
 TGCAACAAGCCCTTCTCAAGATTATTGAGGGAAGTGTGCTAGCGTACCGCCTCAAGGTG
 GACGCAAACATCCACAACAAGAGATGATTCAAGTGGATACAAAAAATATCCTCTTCATCG

TGGGTGGTGCCTTTTGATGGTATTGAAGAAATTGTCAAACAACGTCTGGGTGAAAAAGTCA
 TCGGATTGGTCAAAACAATAAGGCGATTGACGAAAACAGCTCATAATGCAAGAAATCA
 TCGCTGAAGACATTCAAAAATTTGGTATTATCCCTGAGTTGATTGGACGCTTGCCTGTTT
 TTGCGGCTCTTGAGCAATTGACCGTTGATGACTTGGTTCGCATCTTGAAAGAGCCAAGAA
 ATGCCTTGGTGAAACAATACCAAACCTTGCTTTCTTATGATGATGTTGAGTTGGAATTTG
 ACGACGAAGCCCTTCAAGAGATTGCTAATAAAGCAATCGAACGGAAGACAGGGGCGCGTG
 GACTTCGCTCCATCATCGAAGAAACCATGCTAGATGTTATGTTTGAGGTGCCGAGTCAGG
 AAAATGTGAAATTGGTTCGCATCACTAAAGAAACTGTCGATGGAACGGATAAACCGATCC
 TAGAAACAGCCTAGAGGTGACTATGGAACCTTAATACACACAATGCTGAAATCTTGCTCAG
 TGCAGCTAATAAGTCCCACTATCCGCAGGATGAAC TGCCAGAGATTGCCCTAGCAGGGCG
 TTCAAATGTTGGTAAATCCAGCTTTATCAACACTATGTTGAACCGTAAGAATCTCGCTCG
 TACATCAGGAAAACCTGGTAAACCCAGCTCCTGAACCTTTTTTAACATTGATGACAAGAT
 GCGCTTTGTGGATGTGCCTGGTTATGGCTATGCTCGTGTTTCTAAAAAGGAACGTGAAAA
 GTGGGGGTGCATGATTGAGGAGTAATTTAACGACTCGGGAAAATCTCCGTGCGGTTGTCA
 GTCTAGTTGACCTTCGTCATGACCCGTCAGCAGATGATGTGCAGATGTACGAATTTCTCA
 AGTATTATGAGATTCCAGTCATCATTTGTGGCGACCAAGGCGGACAAGATTCCTCGTGGTA
 AATGGAACAAGCATGAATCAGCAATCAAAAAGAAATTAACCTTTGACCCAAGTGACGATT
 TCATCCTCTTTTCATCTGTCAGCAAGGCAGGGATGGATGAGGCTTGGGATGCAATCTTAG
 AAAAAATTGTGAGGAAAAGAAAATGGCAAAAACAATTCATACAGATAAGGCCCCCAAAGGCT
 ATCGGGCCCTATGTTCAAGGAAAAATCGTTGGCAACCTTTTGTGTTGCTAGCGGTCAAGTT
 CCCCTATCCCCTGAACTGGGGAAATTGTAGGAGAGAATATCCAAGAACAGACAGAGCAA
 GTCTTGAAAAACATCGGTGCTATTTTGGCAGAAGCAGGAACAGACTTTGACCATGTTGTC
 AAAACAACCTGTTTCTTGAGCGATATGAACGACTTTGTTCTTTTAATGAGGTTTACCAA
 ACGGCCTTCAAAGAGGAATTCCCAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 797 | 1105 | F | 103 aa |
| 2 | 1179 | 1391 | F | 71 aa |

> 3864458-2 ORF translation from 797-1105, direction F
 VTME LNTHNAEILL SAANKSHYPQDELPEIALAGRSNVGKSSFINTMLNRKNLARTSGKP
 GKTQLLNFFNIDDKMRFVDVPGYGYARVSKKEREKWGCMIEE*

Description:
 unknown

> 3864458-3 ORF translation from 1179-1391, direction F
 VQMYEFLKYIEIPV IIVATKADKIPRGKWNKHESAIAKKLNFDP SDDFILFSSVSKAGMD
 EAWDAILEKL*

Description:

HYPOTHETICAL 22.0 KD PROTEIN IN LON-HEMA INTERGENIC REGION
(ORFX). - BACILLUS S UBTILIS.

Assembly ID: 3864474

Assembly Length: 1673bp

> 3864474 Strep Assembly -- Assembly id#3864474

ACGTTTTGGGAACGTTCGGATAGCAGATTCCGAACAACTGATAATGGTTGGCAAAATC
ATTATTCCTAATAGTAACGAAGCTGGTTAGGACAACATGCCATTTCTAAAAAGGTTT
TAATCCAAGGCACCAATAATTGTAGGCCGAAAAACCATAACAATAGATGGAATGGCTG
CCATCAAGTTGATAGCTGATTTTAAGAAGCTATAGACGGGCTTTGGACAATTATAAACCA
TAAACACCGATGTCAAGATCGCCTGTTGGCACCCCAATCACAATCGCTCCTAAGGTCGAA
TAAATAAGGAACCAACGATCATTGGTAAAATACCATAGCTTGCCGGAATGTTTCGTTGGCG
ACCAATCACTGCCTAATAAAAAACGGGCAAAGCCGTAGTTAGCTATGAAAGGTAAGCCAT
TACTAAAAATAAGAAACAGATTAGCAAAATAGCTACAACAGCTACTGTTGCACTCATGA
AAAAAATTGCCCTAAAACTGCTTCTTTGAAGGCTTGTTTTGTCACATCTTGTCCTTTCT
AGTGAAGAAAGTAAGGGAGATACGACACCTCCCTACTTGCCCTTCTTTATCTTATTGTACG
ATGAAACGTCTGCATCTCTTTAGAGATTTATGGAGCAAACATTTTATTTAATCTTGTCCT
AGGTGGTTAATTTGCCACTAAAAACGTCCGCAAGTTCAGCCATACTGACTTGGCTTGCCCT
TATTGTCATTATTGACCACAACAGCAATACCGTCTAAAGCAATAGCATCATGGGTGAGAC
TCTTACCTTCTTCAGGAGTTAATTCCTTAGAAACCATAACCAATATCAGCGGTTTTCTCCT
TAACAGCGGTAATACCTGCTGAAGACCCATTAGAGGTAATATCAATCGTAACTTCTGGAT
TTTCTTTTTTTATAAGCTTCTGCTAATTTTTTCCATTAAAGAAGATACTGAAGTGGAACCTA
CAACAGACAACCTGCCTGATAAGTGTGGCTTGTTATATTCTGTGGTTTTCGGTTTTAGCTT
CAATAAATTTATTATCTGTGACCACTTGTTGACCTTGTTTGGAGTGGATAAAGCTGATAA
AATCTTGACCTAGCTTGGAAGATTAGAAGACCAACAATGTTGAAGGGACGTTGAAGAG
GGTATTACCATCTAAACTGTGTCTCGACTAGCCTTGACACCATCAATCTCTAAAGCCT
TGACAGATTTTCGTTAAAGATCCCAAGGAGATGTAGCCGATAGCATTAGCATTCCTTGAA
CTGCTGAGAGAACACCTTCTGTACTATTTTGAATCACAGCTGTTTTGGCAGTGTAGTCAA
TTTTTTTTATCACCGTCTTTTTTGAGAATCCCTGTGATTCTGTGAAGGCACCCCGTGTTT
CAGAGCCATTTTCTCGTGAAATCACCTCAATCGTTTCTGGAGCTGACTGTTTGAAGCAG
CTGACTGATTGCCACAGGCAACAAGCCCAAATCCTGATAAGCCAATGGCTGCAAGAGTAA
GCATTTTTTTGAATTTTATAATAATCACCTTTATCTCTATGTATTTTTCTTGTGTAGGCT
TACTACATTTATAGTCTAACAAGTCTTTGTAAAGGTTTATCCCTGATTCATGTAAAGATT
GTGTAAAGAATCAAAAAAAGCCACTTTTGAAAAATGGCTGCCCCCTAAAAATAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 68 | 247 | R | 60 aa |
| 2 | 644 | 1528 | R | 295 aa |

> 3864474-1 ORF translation from 68-247, direction R

VFMVYNCPKPVYSFLKSAINLMAAIPSIYVYGGFGLQLLVPWIKTFLGNGMSCPNQLRYY*

Description:

PROBABLE ABC TRANSPORTER .PERMEASE PROTEIN (ORF72). --BACILLUS
SUBTILIS. (BLAST)

> 3864474-2 ORF translation from 644-1528, direction R
VIIMKFKKMLTLAAIGLSGFLVACGNQSAASKQSAPGTIEVISRENGSGTRGAFTEITG
ILKKDGDKKIDYTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLSKSVKALEIDGVKAS
RDTVL DGEYPLQRPFNIVWSSNLSKLGQDFISFIHSKQGGQVVTDNKFIEAKTETTEYTS
QHLSGKLSVVGSTSVSSLMEKLA EAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMVS
RELTPEEGKSLTHDAIALDGI VVVNNDNKASQVSMAELADVFSGLTTWDDKIK*

Description:

probable hemolysin precursor - Streptococcus agalactiae (strain
74-360)

Assembly ID: 3864510

Assembly Length: 1702bp

> 3864510 Strep Assembly -- Assembly id#3864510
CTTTTTTATTTTACAACAAGTTCATAACGTGTCTTACTGGTGAAGTTTGACCAGCTTTA
AGAATGACTTGGCCTTTAAGGTCACTGTGAATGGCATCTGGTAAAGCTTGCCTTCAAGA
GCAATCCCATTGTGCTGTAGCATTGGCTGACCTCCTATGATGACACTTTTCATCCACAAAG
TTTGCTGTGTAGACCACAAAGCAAGGAGCTTCTGTCTTGAAAAGCAGGAAGCGACCTGAA
TTTTGGTGCATAAAGGAATCCAGCATTGTCATGGCCTGCAGGAAGGGCAAATGGATGATCC
AAACCTGATGCCAGCTGGATTTGCTCATCTTCTTCTGCAAAGATATCCTTCAACAAGGCA
CCATTGTAGATGTGTTTGACCACATCACGGTTGGCTTCTGGAGTTTTTGGCAGGAACACCG
TCAGGAGCGATTGAGTAAATGCCCTCTGTGTTTAGTTGGAAGACATGACGGTCAATCGTC
TGCGTGAAATCACCAGACAAGTTGAAATAGCTGTGGTTGGTTGGATTGACCAGCGTATCC
TGATCGGTCGTTACCTTGTAGATCGAATTCATGGAGGCACCAGTTTCTTCCAAGTGATAA
CTGATCGCCAAATCTTGAGATTTCCAGGGAACCTCCTGTCCCATCTGTACGCTCTGTGT
AGAGAGTCAAGCCATGATCGCTTACTTCTTCAACTTCAAACAAGCTGGAATCCCAACCAG
TTGAACCACTGTGATTACAGTTGCTAGCATTATTAACCTCAAGGTCATAGGTCTTACCAT
TGAGCTCAAAGGTCGCACCTGCAATACGACCCGCTACAGGACCTACACTTGCTCCATGCT
TGGGACTATTGCCTACATAACTATCAAAGTCATCAAATCCCAAGATAACATTGGCAAAAT
TTCCAGCCTTGTGAGGTGCGACATAGCGCAAGATAGTCGCACCATAAGTCATAACCTCAA
GTTGGTAGCCACCGTCTGTCTCAAATCGATAGGCCAAGACATCCTCACCCCTCAACATTTT
CAAATACACGCTCTGTGTATGCTTTCATTCTGTCTCCTTTTACTATTTCTCTCAAGCAA
ACAAACCATAGAAAGCGTACTGACAATCTATGGTTTATCTGATAATTTACAAATCCTCTT
GTCAAGAATTCATAAACA CTGTCTTACTTTTGATATTTCGTGAATTATGACACCTTGTA
ACACGGTTTACTGTACCTGTAGGAGACGGTGTATCTGGTTTATTTTCTACCTTGAGTGAA
GTCAATAGGGCAAAGAGTTGGGCATAAACGATGTAAGGGAAGACACGGTAAATATCATTC

AAGACACCGCCACAACCAAGGGCCACTTCTTTGACATTTTCAAGACCAAAAGCTTGA+CA
 CTCAAAAGCACAAACACGACGAGCAATCTGGTCACCAGCAACTTCACGAACCAAGTCCAAG
 TCGTACTTACGAGTGTAGTCCGTCGTTGTACCAAAGACCAAAACAACACTGTATTGTGCGTTG
 ATAAGAGATTTTGGACCGTGACGGAAGCCAACTGGGCTTTCATACATGGTCGCAACTTGA
 CCAGCAGTTAATTCCAAAATCTTGAGCTGAGCTTCATGAGCAAGTCCAAAGAAAGGACCA
 GCGCCTAGAATAGATGACACGGTTAAAGTCTAAATCAACGAGATCTTTGACATCTTCTGC
 CTTGTCTAAAACTTTACGGGCA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1164 | 1640 | R | 159 aa |

> 3864510-3 ORF translation from 1164-1640, direction R
 VSSILGAGPFFGLAHEAQLKILELTAGQVATMYESPVGFRHGPKSLINDNTVVLVFGTTT
 DYTRKYDLDLVREVAGDQIARRVLLSDQAFGLENVKEVALGCGGVLLNDIYRVFPYIVYA
 QLFALLTSLKVENKPDTPSPGTGTVNRVVQGVIIHEYQK*

Description:

AGAS PROTEIN. - ESCHERICHIA COLI. (Probable tagatose-6-phosphate
 ketose/aldose isomerase)

Assembly ID: 3864526

Assembly Length: 1940bp

> 3864526 Strep Assembly -- Assembly id#3864526
 TGCAGGATTTGATTTGGACGACTTTTATTATTACCAGATTCGCCTAGGAATAGAAAAAAG
 AGCCCAAGAGTTGGACTATGATATCTTGCCTATTTTAAATGACCACCCTTTTACCCTAAG
 CGAGGAAGTGATTGGGATTCTCTGCATCGGAAAGTTTAGTCGAGCTCAGATTTCTGCCTT
 TGAAGAATACCAAAGCCTCTTGTATTTCTAGACAGCGATACACTTCCCTGGGACATAC
 CTGTATTATCACGGATTTTACACTGCTATGAAACAGGTTGTGCGATTATTTCCCTCAGTCA
 AGGAATGGACCGTATCGGGATTCTAACAGGCCTTGAAGAAACAACAGACCAAGAAGAAAT
 CATTCAGGACAAGCGTCTAGAAAACCTTCAAAAACCTACAGTCAAGCGAGGGGAATCTATCA
 TGATGAACTGGTCTTTCAAGGAAGATTTACTGCCCAGTCTGGCTATGACTTAATGAAGGA
 GGCCATTCAGAGCTTGGGAGACCAACTCCGCCAGCATTTTTCGCAGCCAGCGATAGTTT
 AGCTATCGGTGCCCTCCGTGCCCTCCAAGAAGCTGGAATCAGCCTGCCAGATCGCGTCAG
 CCTCATTTCCCTTTAACGACACTAGTCTGACCAAACAGGTCTATCCTCCCCTCTCTAGTAT
 TACAGTTTATACTGAAGAAATGGGCCGAGCAGGTATGGATATTCTTAACAAGGAAGTCCT
 CCACGGTCGGAAAATCCCTAGCCTGACCATGCTGGGAACCAGACTGACATTAAGAGAAAG
 TACCCTAAATCAAGAATAGGATAACATAAAAAACGAATAGAGTTCTAAAACCTCTATTTCG
 TTTTATTATTCGATTACAATCATAGACTTAATGGTCTTACGTTTCATCCATATCTTTGTAGG
 CTTGGTCGATATCTTCCAGTTTATAACTTGAAGTAAAGACGCGACCTGGATTGATATCAC
 CATCAAGGACGGCTTTTAGTAAAAATTGCTTATCGTATGTTGTAGCAGAAGCTGCCCCAC

CTGCTACAGAGATATTTTGCATAAATGTGCAACCAAGAGCACGATTATTATAGTGTGGGA
 CTCCTACAAAGCCCATACGCCCTCCATTATGAAGAACACCTAGCGCCTGTTCTATAGCAG
 CCTCCGTACCAACACATTCAAGTGCTGCGTCTGCTCCTCCGCCGAGGATTTACGCACCT
 TGGTAATTCCTTCTTGACCACGTTCTGCAACAACAGCTGTGCGACCTGACTCCATAGCCA
 TCTTTTGACGGTCTTCATGACGGCTCATAAGGATAATTTGTGATGCTCCACGCATCTTAG
 CCGCGATGACAGCACATTGACCAACAGCCCCATCACCGATAACAACAACCTTGTCCCCTT
 TTTGAACATTTGCAACACGCGCCGCATGATAGCCTGTGCGCATGACATCTGCAAGAGTCA
 AAAGGGACTTGAGCATCCCTTCTGTATAGTCAGAAGGTTGACCAGGGATTTTAACCAGCG
 CCCAGTTTGCATAGTGGAAGCGAATATATTCTGCCTGAAAATCACCCCCCAAATTATTGC
 CAATATGATTGTGCGCAAGAACCGTCAAATCCAGCAAGACAGGCATCACACTCACCACATC
 CATGTGTAAAAGGGACAATCACAAAATCACCTGGTTTCACCGTCGTAATGGCTTCCCCAG
 CTTCTTCAACAATCCCAATCGCTTCGTGTCCACTTATTTTTTGTGTCCAACCTTCGTTTT
 CCNTGGATTACGGTACCTCCATAAATTTGAACCACAAACGCACGCACGAACCACACGAAT
 AATCACATCATCCGCTTCTATTATTTGCGGACGTTCAATGCTAGCAAGTCCAACCTGACC
 TGCCTTTGTATATACTGCTGATTTCAATTTAAATTTTCCTTCCTTATAAAGTTTAATTTT
 GAGATTTAAACGATTTAAAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 845 | 1660 | R | 272 aa |

> 3864526-2 ORF translation from 845-1660, direction R
 VKPGDFVIVPFTHGCGECDACLAGFDGSCDNHIGNNLGGDFQAEYIRFHYANWALVKIPG
 QPSDYTEGMLKSLTLADVMPTGYHAARVANVQKGDKVVVIGDGAVGQCAVIAAKMRGAS
 QIILMSRHEDRQKMAMESGATAVVAERGQEGITKVREILGGGADAALCEVGTAAIEQAL
 GVLHNGGRMGFVGVPHYNNRNLGSTFMQNISVAGGAASATTYDKQFLLKAVLDGDINPGR
 VFTSSYKLEDIDQAYKMDERKTIKSMIVIE*

Description:

ALCOHOL DEHYDROGENASE (EC 1.1.1.1). - ALCALIGENES EUTROPHUS.

Assembly ID: 3864548

Assembly Length: 2051bp

> 3864548 Strep Assembly -- Assembly id#3864548

ATCGAATTTTTCTAGCCAGGCTACAGTTTTGGCAAGTAAGGTTTCATCTCAGGCAGTCAA
 CTGGGTGAGTGCCCTTTATTAGCGGAGCTTCTCAAGTGATTGTTGCCTTGATTATCGTTCC
 TTTCATGCTCTTTTATCTCTTGCGTGATGGGAAAGGCTTGCCTAACTATTTGACCCAATT
 CATTCCAAGAAAAATTGAAGGAACCTGTTGGACAAGTTCTATCAGATGTGAATCAACAGTT
 GTCCAACATATGTTGAGGGCAAGTGACAGTGGCTATTATTGTAGCAGTAATGTTTATCAT
 CTTCTTCAAGATTATTGGTCTACGCTATGCGGTTACGCTGGGGGTTACTGCTGGTATTTT
 AAATCTGGTCCCTTATCTTGGTAGCTTTCTAGCCATGCTTCCTGCCCTAGTATTGGGTTT

GATTGCTGGTCCAGTCATGCTTTTGAAGTAGTGATTGTCTTTATTGTAGAACAAACTAT
TGAAGGCCGTTTTGTCTCTCCATTGATTTTGGGAAGTCAATTAAACATCCACCCTATTAA
TGTCTCTTTGTTTTGTAACTTCAGGATCTATGTTTGGTATCTGGGGAGTTTTACTTGG
TATTCGGTTTTATGCCTCTGCTAAGGTTGTCATTTTCAGCCATTTTCGAATGGTATAAGGT
AGTCAGTGGTCTATATGAATTAGAGGGTGAGGAAGTCAAGAGTGAACAATAGTCAACAGA
TGTTACAGGCTTTGGAGGAGCAAGATTTAACTAAGGCTGAGCATTATTTGCCAAAGCTT
TAGAAAATGATTCAAGTGATCTTCTGTATGAGTTGGCAACTTATCTTGAAGGGATTGGTT
TCTATCCTCAGGCCAAGGAAATTTACCTGAAAATTGTAGAAGAATTTCCAGAGGTTTCATC
TTAATCTAGCTGCAATGGCTAGCGAGGATGGTCAAATAGAAAAAGCCTTTAACTATCTTG
AGGAAATCCAAGCTGACAGTGACTGGTATGTCTCGCTCTTTGGCTCTGAAGGCAGACCTA
TACCAGCTGGAAGGTTTGACAGATGTGGCACGTGAGAAATTATTGGAGGCCTTGACCTAC
TCAAAGGATTCTCTCTTGATATTGGGTTTGGCAAAGTTGGATAGTGAGTTGGAAAAATTAC
CAAGCGGCTATTCAAGCCTATGCCCAGTTAGATAATCGCTCGATTTATGAGCAAACGGGC
ATTTCCACCTATCAACGAATTGGCTTTGCCTATGCTCAGTTAGGGAAATTTGAAACGGCT
ACTGAGTTTTTTAGAAAAAGCCCTGGAGTTAGAATACGATGACTTAACAGCTTTTGAGTTG
GCCAGTCTTTATTTTGATCAAGAAGAATATCAAAAAGCCACCCTCTACTTTAAGCAGCTT
GATACCATTTCTCCTGACTTTGAAGGCTATGAGTATGGGTACAGTCAGGCTTTACATAAG
GAACATCAAGTTCAAGAAGCCCTGCGTATCGCTAAGCAAGGATTAGAGAAAAATCCCTTT
GAAACTCGCCTCTTGCTAGCTGCTTCACAATTTTCTTATGAATTGCATGATGCTAGTGGT
GCAGAAAATTATCTCCTTACTGCAAAAGAAGACGCTGAGGATACAGAAGAAATCTTGCTT
CGTTTAGCCACTATTTATCTGGAGCAGGAGCGTTATGAGGATATTCTAGACTTGCAGAGT
GAGGAGCCAGAAAATCTTTTGACCAAGTGGATGATTGCTCGTTCTTATCAAGAAATGGAC
GATTTGGATACTGCTTATGAGCATTATCAAGAGTTGACAGGAGATTTGAAGGACAATCCA
GAATTTCTGGAACACTATATCTATCTCTTGCGTGAATTGGGACATTTTGAAGAAGCAAAA
GTCCATGCTCACACTTACTTAAACTGGTTCCAGATGATGTGCAAATGCAAGAACTGTTT
GAGAGATTGTAAGAATGTTTAAACATATAGAAGTGTAGTTTATCTCTTTTGATAGCTACG
GTCTTTATTTGTACATGGTAGAATCTTTTACAAAATACTTGGTAATCTTGTTTATTCA
TGCCATAATAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 687 | 1055 | F | 123 aa |
| 2 | 979 | 1932 | F | 318 aa |

> 3864548-2 ORF translation from 687-1055, direction F
VRKSRVNSQQMLQALEEQDLTKAEHYFAKALENDSSDLLYELATYLEGIGFYPPQAKEIY
LKIVEEFPEVHLNLAAMASEDGQIEKAFNYLEEIQADSDWYVSLFGSEGRPIIPAGRFDR
GT*

Description:

unknown

> 3864548-3 ORF translation from 979-1932, direction F
 VTGMSRSLALKADLYQLEGLTDVAREKLLEALTYSKDSLILGLAKLDSELENYQAAIQ
 YAQLDNRSIYEQTGISTYQRIGFAYAQLGKFETATEFLEKALELEYDDLTA FELASLYFD
 QEEYQKATLYFKQLDTISPDFEGYEYGYSQLHKEHQVQEALRIAKQGLEKNPFETRLLL
 AASQFSYELHDASGAENYLLTAKEDAEDTEEILLRLATIYLEQERYEDILDLOSEEPENL
 LTKWMIARSYQEMDDLDTAYEHYQELTGDLKDNPEFLEHYIYLLRELGHFEEAKVHAHTY
 LKLV PDDVQMQLFERL*

Description:

unknown

Assembly ID: 3864582

Assembly Length: 1318bp

> 3864582 Strep Assembly -- Assembly id#3864582

CTTTAGCAATCAGTTTATTGGGAGATTTGACTGCCACTTCTGTTGGAACCTTGATAATCT
 TTTTACCCTCAAAGCGTTCCATACCAGAAATCTTAACATCAACTGCTAAAATAACTACAT
 CCGCTGCATCAATCTGCTCTTGACTCAATTCATTTTCTACCCCTATTGTCCCCTGAGTCT
 CAACATGAATCACATGTCCAGCTACCTTTGCGGCATTCTCTAATTTTCTGTGCAATAT
 AAGTGTGGGCAATTCCCATAGTACAAGCTGCAACACCAACAATTTTCATACGGATACCCT
 CCAAAATTTTTTCTTATTAACAAAAAGCTGCAATCACATCATCAGATGTCTGAGCCCGAA
 CTAATTTGGCAACAACCTTCGTCATTACCAAGTTTTTCGAGCAAAGAGTGATAAGGTCTTCA
 AATGCTCCCTAGCAGCTTCTGTATCATCACCAACTGCAAAGAGTACAATTACTTTGACCC
 CTTTCCCATCAATGGTCTCCCAAGGAATCTCATTGTGATTTATAGCTATGACTACCCCG
 CCTTCTCCACAGCAGAACTCTAGCTATGGGGAATAGCAATATAATTTCCAATACCGGTCT
 GTCCTTCTGCCTCTCTCTGATAAAGACCTTCGATAAATTGGTCTCTATCAGACACATAAC
 CCGTCTCAACCAATAGTATGAGCTAATGCCTCAAAAACCTCTTCTTTGCTCTGCATCTGT
 AAATCCGTCTGGATCAGACTCACATTAAGAATATCTTTGATTTCCATATATTATCTCCCG
 TAATTCCTTCTTTGTTAACTGTTTTAATTGATTTATGAATGATTCATCTGCTAGTCTTCT
 CATCAATGTTTTAATACATGACTTGTCTGTGATACTGCAATGGCCAAACCGATAATAAG
 GTCAACACACTGGATATCCTTCGACCATCTCTGATAGGTGGTTTTAATCTAGTAATCAC
 TAAGACATGATGTTGAAAGTTTCCTTCACAATGTGGTAGAAGAACACCTTTAGCAACCTC
 TATACTTCCCTGTCTCTCACGGTAATATAGAAGCTCTTCTATTTTTTCTGTATCTTCAGA
 AACAAGAAGGCTGATTTGATTTGCTAATTCCTTTGTAGGCTTCTTGACGATTTTGAACAGA
 TATATCCATAAGGACAAGCGAAAGATTATTCATAGTTTATCTCCTGAATTTTTGCTTGAA
 GACGTTGTTTATCACCTCGGTTAGAAAAGCACTAACTAGGACAAACGGGACACTTGCTG
 GTTCCTGCAAAGCTACCGTCGTCACAATGAAATCTAAATCTGGATATAGATTTATCAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 317 | 550 | R | 78 aa |

> 3864582-1 ORF translation from 317-550, direction R
VEKAGVVIAINHNEIPWETIDGKGVKVIVLFAVGDDTEAAREHLKTLSLFARKLGNDEVV
AKLVRAQTSDDVIAAFC*

Description:

Probable phosphotransferase enzyme IIa component

Assembly ID: 3864604

Assembly Length: 2077bp

> 3864604 Strep Assembly -- Assembly id#3864604

CTAGTCTTGGCTACTGTCTAAGTTGGCTTGTGCATAAGCCTGCCAGATTTTTTGTGGGG
TTTGGCAAGTGGGTAATTCTTGAATTCTTCTGGTGAAAGCCAACGAACCTCCCTATCTGA
AAAATCATGGAAGTCACTCACCTGACCTGCTACAATCTGTACATGCCATTTTCGATGACT
AAAAACATGCTGGACTGTATCAAAACAAACATCAAGCCAATCAACATCTAGGTCATAGTC
CTGCTGGAAACTCTTCTTGGGACTGGGGCCAGAGTTCACACTTTCTTCCGCAACCTGAT
GAAAGAGGTCAAACCTGCTCTTCTTGCGAAAAGTTATCAACTTCTATAAAGGGGAAATGCC
AAAAACCTGCCAAGAGCTTTTCGCTTTTCATTTTTTTTCAAGTAAAAATTGTCCTTGAGAAT
TTTTTCACAACATAAGGCTTTAAGATAAATAGGAACCGGCTTTTCTTAGGAGATTTAATTG
GATAACGGTCCATGGTTCATTCTGATATGCCGCACTAAAGTCCTTGACTGGGCTTTCTT
CAGGTCTGGGATTTACAGGAGACTCAATATCAGACCCTAAGTCCATCAAGGCTTGATTAA
AATCACCCGGACGATCTGGATTAATCAAGATCTCCATCATTGCCTGAAAAATTTTTCGAT
TACTTGGAATCCCAATATCGTGGTTGACTTCAAACAGACGCGCCAAGACCCGCATGACAT
TACCATCTACAGCTGGCTCAGGCAAGTTAAAAGCAATACTGGAAATGGCTCCTGCTGTGT
AAGGTCCAATCCCTTTCAAGCTGGAAATTCCTTCATAGGTATTTGGAAATGGCCACCAA
AGTCAGTCATAATCTGCTGGGCTGCAGCCTGCATATTGCGAACTCGAGAATAATAACCCA
AGCCCTCCCAAGCTTTTCAGTAAACTCTCCTCAGGCGCAGTTGCCAGACTTTCGACAGTTG
GAAACCAGTCCAAAAATCTTTCGTAGTAAGGGATAACTGTATCCACCCTGGTCTGCTGAA
GCATGATTTTCAGATACCCAGATGTGATAAGGATTTTTTACTTCTCCTCCAAGGCAAATCTC
TTTTGTTTTTTCATCATAACCAAGCGAGAAGTTTTTCTCACCGGAAAGAAATGACTTTCTCCTC
CGGCCACATGACGATACCGTATTCTTTCAAATCCTAACATATCTCTAGTTATAACACAGA
AGGTTTCACCTGTCTTTGTATCTGATTTATAATATTTTTCAATAGATAGTATATAACTTTT
CCTATCTACTTATACTCCAATGAAAATCCAAAGAGCAAACCTAAGAAGCTAGCCGCAGGTT
GCTCAAACACTGTTTTGAGGTTGTGGATAGAAGTACAGAGTCAGTATCATATTACCTA
CGGCAAGGTGAAGCTGACGTAGTTTGAAAAGATTTTTCGAAGAGTATAAATCTTATTGATG
AACTGCTTGCACTCTGAGAAAAAATGAGCTTGGATATTATTCCAACTCACTTAAAGTC
AATTTCAATCCACTAGAACAAAGCCTAGTACAGTTCCATCGCTTTCAACATCCATGTTGAG
AGCTGCTGGACGTTTTTGAAGACCTGGCATGGTCATAACATCACCAGTTAAGGCAACGAT
GAAGCCTGCACCTAATTTTGGTACCAATTCACGAATGGTAATTTCAAAGTTTTCTGGTGC
TCCAAGCGCATTTGGATTGTCTGAGAACTGTATTGAGTTTTAGCCATACAAATTGGCAA
TTTGTCCCAACCGTTTTGAACGATTTGAGCAATTTGTGTTGAGCTTTCTTCTCAAAGTT
CACTTTGCTACCACGATAGATTTTCAGTGACAATTTTTTCAATCTTTTCTTGGACAGAAAG
GTCATTATCGTACAAACGTTTATAGTTAGCTGGATTTTCAGCAATTGTCTTAACAACCTGT

TTCGGCAAGTGCTACTCCACCTTCTGCTCCATCAGCCCAGACACTAGCCAATTCAACTGG
TACATCGATTGAGGCACAGAGTTCTTTTAAGGCTGCAATTTTCAGCTTCTGTATCAGATAC
AAATTCGTTAATAGATACAAGCTAATGGAATACCGAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1 | 141 | R | 47 aa |
| 2 | 1513 | 1803 | R | 97 aa |

> 3864604-1 ORF translation from 1-141, direction R
VSDFHDFSDREVRWLSPEEFKNYPLAKPQQKIWQAYAQANLDSSQD*

Description:
unknown

> 3864604-3 ORF translation from 1513-1803, direction R
VNFEKKAQTQIAQIVQNGWDKLPICMAKTQYSFSDNPALGAPENFEITIRELVPKLGAG
FIVALTGDMVTMPGLPKRPAALNMDVESDGTVLGLF*

Description:
FORMATE--TETRAHYDROFOLATE LIGASE (EC 6.3.4.3)
(FORMYLTETRAHYDROFOLATE SYNTHETAS E) (FHS) (FTHFS). -
CLOSTRIDIUM ACIDI-URICI.

Assembly ID: 3864610
Assembly Length: 1887bp

> 3864610 Strep Assembly -- Assembly id#3864610
CTCAAAACNCTGCTTTGAAGAGATTTTCAAAGAGTACAAGAAGTTTAGTTATTAGCGTTC
TTACCGCTTGTAAGACTAGATTTCTCATAAAATAGAATCTTTTCCTTTTAGTTGTAAACTA
GTCTGGGAGAGTAGAGAGGTTTGAGATACCTTTCTAGCTTTTGGATTATCATCTAAGAAG
AGTAATTTCCCTTGCAATTAAGAGGGGAAAAAGAGACACGAAATGACTATAATGGGTGAC
AATGGGGGAAGGGATAGACAAGAGATTTTATCCACATATGAAAAAGGAGGTTAGGAAAG
AGTTATATATCCTATATTATATAAATAATCAATTGCGCAGAAATTTGGTAAGAATTCATG
CGTCAACTCATAAAGAACTACTTAAAAAATTCACAGTATTCATAATTATTTTCGAGGAGA
AAAACAGTGAAAAAAGAAAAAGCTTGCTCTGTCTCTTATCGCTTTTGGCTGACGGCT
TGTTTAGTAGGCTGTGCTAGCTGGATTGATCGTGGAGAATCCATAACGGCTGTTGGCTCA
ACTGCCTTGCAACCCTTGTTGAAGTAGCGGCAGATGAATTTGGCACCATCCATGTTGGA
AAAACGGTCAATGTCCAAGGGGAAGTTCTGGTACAGGCTTGTCACAGGTTGAGTCTGGG
GCAGTTGATATAGGAACTCAGATGTATTTGCTGAGGAAAAAGACGGAATTGATGCTTCT
GCTCTTGTTGACCACAAGGTCGCGGTAGCTGGCTTGCTCTGATTGTCAATAAGGAGGTT
GATGTTGATAACCTAACGACAGAGCAACTTCGTCAAATCTTCATAGGTGAGGTAACCAAT

TGGAAGAGGTTGGTGGTAAGGACTTACCCATCTCTGTTATCAATCGGGCAGCCGGCTCT
 GGCTCTCGTGCTACCTTTGATACTGTCAATTATGGAAGGTCAGTCTGCCATGCAAAGTCAG
 GAGCAGGATTCAAATGGAGCGGTAAATCAATCGTATCAAAAAGTCCAGGAGCTATCTCT
 TATTTATCTCTTACCTATATAGATGATTTCGGTCAAAAAGCATGAAGTTGAATGGCTATGAC
 TTAAGTCCAGAAAATATAAGTAGCAATAATTGGCCCTTGTGGTCTTATGAGCATATGTAT
 ACATTGGGGCAGCCCAATGAGTTGGCTGCAGAATTTCTCAATTTTGTCTCTCGGATGAG
 ACCCAAGAAGGGATTGTCAAAGGATTGAAGTATATTCCGATTAAGGAAATGAAGGTTGAA
 AAAGATGCTGCCGGAAGTGTGACAGTGTGGAAGGGAGACAATAATGAATCAAGAAGAAT
 TAGCTAAGAAAATGTTGCTTCCATCAAAGAATTCTCGTCTGGAGAAAATTAGGAAAAGGTT
 TGACCTTTGCCTGTCTTTCTTTGATAGTCATCCTTGTGGCCATGATTTTGGTTTTCGTAG
 CGCAAAAAGGCTTGTGACCTTCTTTGTCAATGGTGTGAATATCTTTGACTTTCTTTTGG
 GAGGAAGTTGGAATCCTTCTAGTAAAGAAATTTGGTGGCCCTTCCATGATTTTGGGTTCCCT
 TTATCGTTACCATTCTCTCAGCCCTTATCGCAACACCCTTTGCTATTGGTGCAGCAGTTT
 TTATGACCGAAGTATCACCAAAAGGGGCGAAGATTTTGCAACCAGCTATTGAAGTCCCTGG
 TTGGGATTCCCTTCAGTAGTGTACGGATTTATTGGCTTGCAAGTCGTCGTTCCCTTTGTTTC
 GCAGTGTCTTTGGTGGGACTGGTTTTGGGATTTTGTGAGGATTTCCGTCCTCTTTGTCA
 TGATTTTGCCGACCGTAACCTTTATGACAACGGATAGCTTGCGTGCGGTTCCCTCCNTTAT
 TATCGTGAAGCCAGTTTCGCTATGGGA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 427 | 1305 | F | 293 aa |

> 3864610-1 ORF translation from 427-1305, direction F
 VKKRKKLALSLIAFWLTACLVGCSWIDRGESITAVGSTALQPLVEVAADFGTIHVGKT
 VNVQGGSSGTGLSQVQSGAVDIGNSDVFAEEKDGIDASALVDHKVAVAGLALIVNKEVDV
 DNLTTQEQLRQIFIGEVTNWKEVGKDLPISVINRAAGSGSRATFDTVIMEGQSAMQSQEQ
 DSNQAVKSIIVSKSPGAISYLSLTYIDDSVKSMKLNQYDLSPENISSNNWPLWSYEHMYTL
 GQPNELAAEFLNFVLSDETQEGIVKGLKYIPIKEMKVEKDAAGTQTVTVLEGRQ*

Description:

PROBABLE ABC TRANSPORTER BINDING PROTEIN PRECURSOR (ORF108). -
 BACILLUS SUBTILIS. (BLAST)

Assembly ID: 3864716

Assembly Length: 405bp

> 3864716 Strep Assembly -- Assembly id#3864716
 CTGAGGAATCAAAAGTTGAACCACCAGTAGAACAAGCATAAGTCCCAGAACAACCCGTGC
 AACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACAACAAGAAAATCCTA
 AAGAAGATAGGGGAGCGGAAGAGACTCCGAAACAAGAAGATGAACAGCCAGCAGAAGCCC
 AAGAAATCAAGGTTGAAGAACCAGTAGAATCTATAGAGGAGACTGTCATTCAACCTGTTG

AACAACCAAAAGTGGAAACGCCTGCTGTTTAATAACTAACGGAACCTACAGAGGAACCTA
AAGTTGAAGTAACTAGTATTCCCCTCACTACTCGCTATGAGGAAGACCTTACTTACGAAC
ACGGAACGCGTTGAAGTTGTTAAGGAAGGTTATAATTGGCAGTAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 57 | 272 | F | 72 aa |

> 3864716-1 ORF translation from 57-272, direction F
VQPTQAEQPSTPKESSQENPKEDRGAEETPKQEDEQPAAEQEIKVEEPVESIEETVIQPV
VEQPKVETPAV*

Description:

unknown

Assembly ID: 3864718

Assembly Length: 1542bp

> 3864718 Strep Assembly -- Assembly id#3864718

CTATGGGATTGGTAGTTCTTCTTAGTGCAGGGGCTGTAGACCCAGTTGCGACCCTAGCGC
TGGACTAGTCGAGAGGGTGTGTTGAAAATGGATGGCTATCGCTATGTTGGTTATCTATC
AGGTGACATCCTCAAACGCTTGGCTTGGACACTGTTTTAGAAAGAAACCTCAGCAAAACC
TGGAGAGGTGACTGTAGTCGAAGTTGAGACTCCTCAATCAACAACAAATCAGGAGCAAGC
TAGGACAGAAAACCAAGTAGTAGAGACAGAGGAAGCTCCAAAAGAAGAAGCACCTAAAAC
AGAAGAAAGTCCAAAGGAAGAACCAAAATCGGAGGTAAAACCTACTGACGACACCCTTCC
TAAAGTAGAAGAGGGGAAAGAAGATTACAGCAGAACCATCTCCAGTTGAAGAAGTAGGTGG
AGAAGTTGAGTCAAAACCAGAGGAAAAGTAGCAGTTAAGCCAGAAAGTCAACCATCAGA
CAAACCAGCTGAGGAATCAAAGTTGAACCACCAGTAGAACAAGCAAAAGTCCCAGAACA
ACCCGTGCAACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACAACAAGA
AAATCCTAAAGAAGATAGGGGAGCGGAAGAGACACCGAAACAAGAAGATGAACAGCCAGC
AGAAGCCCAAGAAATCAAGGTGAAGAACCAGTAGAATCAAAGAGGAGACTGTTAATCA
ACCTGTTGAACAACCAAAAGTGGAACGCCTGCTGTAGAAAAACAAACGGAACCAACAGA
GGAACCAAAAGTTGAAGTAACAAGTATTCCCCAACTACTCGCTATGAGGAAGACCTTAC
TAAGGAACACGGAACGCGTGAAGTTGTTAAGGAAGGTAAGAATGGCAGTAGAACAGTTAC
TACTCCATATATCTTGAATGCGACAGATGGTACGACTACAGAAGGCACCTTCGACAACCTGA
TGAAGCTGAGATGGAGAAAGAGGTTGTTTCGTGTTGGCACGAAACCCAAAGAAAAATTAGC
TCCAGTCTTAAGTTTGACAAGTGTTACAGATAATGCAATGTTGCGTAGTGCGAGACTTAC
TTATCATTTTGAAAATACAGATAGTGTTGATGTGAAAAAAATTCATGCTGAAATTAAAAA
TGGCGATAAGGTTGTCAAACCTATTGACTTATCTAAAGAGAGATTATCAGATGCTGTTGA
CGGTCTTGAACCTTTATAAAGATTATAAGATTGTGACGAGTATGACCTATGATAGAGGTAA
TGGTGAAGAAACCTCTACGTTGGAAGAACTCCACTACGATTAGACCTCAAGAAGGTTGA
ATTGAAAAACATCGGCTCTACTAATCTCGTCAAAGTAAATGAGGATGGTACTGAGGTGGC

AAGTGA~~CTTCTTAACAAGTAAACCTGTGGATGTGCAGAATTACTACCTCAAAGTAACTTC~~
 CCGTGATAATAAAGTTGTTTCCCCTCCCAGTTGAAAAAATTGAAGAGGTGACTGAGGAAG
 GTCCACCACTTTACAAAGTCCCTGCTAAGGCCCTAATTTGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 77 | 1474 | F | 466 aa |

> 3864718-1 ORF translation from 77-1474, direction F
 VLLKMDGYRYVGYLSGDILKTLGLDTVLEETSAKPGEVTVVEVETPQSTTNQEQARTENQ
 VVETEEAPKEEAPKTEESPKKEPKSEVKPTDDTLPKVEEGKEDSAEPSVVEEVGGEVESK
 PEEKVAVKPESQPSDKPAEESKVEPPVEQAKVPEQPVQPTQAEQPSTPKESSQOENPKED
 RGAETPKQEDQPAEAQEIKVVEEPVESKEETVNQPVQPKVETPAVEKQTEPTEEPKVE
 VTSIPQTTTRYEDLTKEHGTREVVKEGKNGSRTVTTPIYILNATDGTTEGTSTTDEAEME
 KEVVRVGTKPKEKLAPVLSLTSVTDNAMLRSARLTYHLENTDSVDVKKIHAIEIKNGDKVV
 KTIDLSKERLSDAVDGLELYKDYKIVTSMYDRNGEETSTLEETPLRLDLKKVELKNIG
 STNLVKVNEGDGTEVASDFLTSPKVDVQNYLKVTSRDNKVVSPPS*

Description:

unknown

Assembly ID: 3864802

Assembly Length: 1321bp

> 3864802 Strep Assembly -- Assembly id#3864802
 ATCGAATTACTTCAACTCCAACCTTACTCTCAATAAAAATCAAATGTAAAAAGAGGAGCT
 AAATTTATCTTTTTCTCCTCCTTCATCGTTCTTACTTTTGACCATAATAAGCATTGCTC
 CATGTTTACGTTGGTAGTGTTTTCTAGTATGTACTGGGGAGCAGGTTCAACTCTTGAT
 TGATTTGTTCTGTAAAGCGATTCATCTTTGATACTTCCTCTAGTACGACAGAGTGATAAA
 CAGCATTCTCTGGATTTTTGCCCCAGGTGAATGGACCGTGATTGCGTACAACAATTCCTG
 GTACTTCAACCGGGTTAAGTCCGCGATGTTCAAACCTCTTCTACGATAACCAGGCCAGTAT
 CTTTTTCATAGGCCACTTCTACTTCGTCCTTGGTCAAACCTACGGGCGCAAGGGATTGAAC
 CGTAGAAATAATCTGCATGGGTGTTCCGTAGAAAGGAATATCACGACCTGCCTGAGCCC
 AAGCAACAGCTTCTGTGCAATGGGTGTGAACCACACTACCAATTTCTGACCAAGCCTTAT
 ATAATTGCACATGAGTTGGGAAGTCGGAAGATGGTCTTAAATCCCCCTTATAGGATCTTAC
 CATCTAGATCAGTCACTACCATGTTTTTCAGGTGTCAATTCGTCATAATCCACGCCTGATG
 GTTTGATAACAATGACACCGAGTTCGCGATTGACTTCAGATACATCCCCCAGGTAAATT
 TGACAAGTCCATGTTTTGGCAATGATTGATTGGCATCACAGACTCGTTTTACGCATAGCAT
 TGATTACTTGATTTCATCTTACATCAAACCTGCTTCTTAATGAGTGGATAGAGAAAAGCT
 TGCGCCTCTTGAATGGCTGCGCGTGTTTCTTCTACTGTTTCAAAATTTTTCAGACCACATT
 TCGATTAGGAAAGGTCCATTATAATTGGTTTCCTTTAAAATATCGAAAGCTTCTTCCCAT
 TTGACACAACCTTGCCCAAAGGTACATCTCGGAACCTGGCCCTTTGAACTTTCTGTCACT

GCATAAGTATCCTTGAGATGGAGAGTTGCGATGGCATGATGACCAAGATAAACTCACTA
 TAGATATCATTATGCCATGCAGACACATTACCAATATCTGGATATACAAAGAGGAAGGGA
 GAGTCAATCTCTTTTCTATAGCCAAATATTTTTCGATGCTATTGATGAAAGGATCATCC
 ATAATTTCAATAGCAAGTACCACCTGAGCTTCTTCAGCCCAGTCACAGGCTTTTCTCAA
 TTTTGTATAAAACGTTGGCGTGTCTGGGGTGACTTTTCCTCATAGTAAACATCGTAACCA
 G

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 92 | 550 | R | 153 aa |

> 3864802-1 ORF translation from 92-550, direction R
 VQLYKAWSEIGSVVHSTHSTEAVAWAQAGRDIPFYGTTHADYFYGSIPCARSLTKDEVEVA
 YEKDTGLVIVEEFEHRGLNPVEVPGIVVRNHGPFWTGKNPENAVYHSVVLEEVSKMNRFT
 EQINPRVEPAPQYILEKHYQRKHGPNAYYGQK*

Description:

L-RIBULOSE-5-PHOSPHATE 4-EPIMERASE (EC 5.1.3.4). - ESCHERICHIA
 COLI.

Assembly ID: 3864854

Assembly Length: 1265bp

> 3864854 Strep Assembly -- Assembly id#3864854
 TTTTCTGTTTTTCGGAGCAAACCTGGGCTCCAGCCGGTTTTGGCCTTCTTTTCCTTAGCTA
 CAGCTGGTTTTAGCTGGCTCAGATTTTTTCGGCTTCTTTTCTGCACTTACTTTTGGTGCTG
 CAGGTTTTGCTTCTACTTTTCGGAGCAGCTGCAGGCTTAAAGCTGGCAGCAATTTTTGCAG
 CGACAGCTTCTTCCACACTTGATGAGTGGCTTTTCACATCCAAGCCCAACTCTTTTGCAC
 GCGCTACAACCTTCTTTACTTTCTTTTCCAAGTTCTTTTTCGATTTCGTACAATCTTTTCT
 TAGACAAATCATGTCCTCCTCTTCTATTCCATAAGAGACCTCATTTTCTTTGTAAATCCA
 GCATCTGTTACAGCCAAAACCTTTCTCGATTTCCCGACTGCTATGATTAATTCCAGTGTT
 GAAAACACGGTTACAATTTCTACTTGATAATAATGACTTTTATCTTGAATCTTCTTGCTC
 AGATTGGGTCCAGCATCATGAGCTAGAAAGACCAACTTGCCCTTGCCGTCTTGAATGGCC
 TTGACCACCAATTCTTCACCCGATATGATGCGCCCTGCTCGCTGAGCAAGCCCCAAGAGA
 TTACTTATCTTTTGCTTATTCAAGTCCCAACTCTCTTCTTTTCACTTTGTGATCCACATA
 AGCGATCAACTCGTCATAAAAGCTTTCTTCCACTTCCATGCTAAAGCTGCGGTAAAGAC
 CTTCTTCTTTTTCGCCTCTAGGGCTTCTGCATTGTCTAGTTTGATATAAGCGCCGCGCC
 ATTGGCCTTGCCCGTAGGATCAATAAAGACTTGTCCTTCCTTGTCTTGACAATGCGGAG
 CAAATCACGCTTATCAATCACTTCGTTAGACACAACAGACTTGCGCAAAGGGATTTTTCT
 TGTTTTCATCTTTCCCTCCTCTAGCAGCTTTTATTCTTCTACAGTATCGTTTTCTACTTC
 CAACTCTACTGAAGCAGCGTCTTCCATGGCTTCAAATTCGCTAGCAGACTTGATATCGAT
 ACGGTAACCAGTCAAGTGAGCCGCCAAGCGCACGTTTTGTCCACGACCAATGGCAAG

AGAAAGCTTGTTATCTGGAACAACCACCAAGGCACGTTTGCTGTCGTTTTTCATCAAAGAT
AACTTGGTCAACCTCAGCAGGAGCGATGGCATTGTAGATAAATTCAGCTGGATCTGCTAC
CCACTCGATAACATCGATATTTTCTTCGATTGGTACCATGCGGTCATTTTTAGCATCGTA
ACGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 324 | 548 | R | 75 aa |

> 3864854-1 ORF translation from 324-548, direction R
VVKAIQDGKAKLVFLAHDAGPNLTKKIQDKSHYYQVEIVTVFSTLELIIAVGKSRKVLAV
TDAGFTKKMRSIME*

Description:

PROBABLE 11.1 KD RIBOSOMAL PROTEIN IN NUSA-INF B INTERGENIC
REGION (ORF4). - BACILLUS SUBTILIS.

Assembly ID: 3864862

Assembly Length: 1305bp

> 3864862 Strep Assembly -- Assembly id#3864862
ATAAACCAAGGAAGCTGAGCTCTTTAGTCCCAGCTTCTTTTTATATATAAAATTTTACC
CGTGAAAAGACAGGGCCTTAGCAGACTTCTTTTTTACTTCGTTACCCCTTGCTTTTTCTT
TGTATGTTTGGGCGTTGGCAGTTGGTTATACATAGCTAAAATCAGGTCTTATAGAAACAT
CTTATTATCAAGTTCTTCCACTCAAATCATTTCTTTGGCACCTTTGTATGGAAACTCAA
AGAAGATTGGTCAATCTTATCTAAGACTGCTTGCACGGGTTTAACTAAAAGCGATCGTCA
TAAATGCCGCCAATAATCTTGCCGCGGAAGTAAAGAATATACTCCCCCATCATGGAACGG
TAAGTCACATCATCTAATCCTGATAATTGTTCCAAAACAAATTCCAAATAGTTCTTACTT
GATGCCATTTCTAATCTTCTAGGCTCTGTTCAACGATAACAACCGTATAGAGTTCTTGCT
TAACCTCGCATCCAATTGATTTAAAGCCCTGCTTTTCCAAAAATGCTGAGATTGCGGAT
TTCCCTTAACATAAGCCAAACGTGCCTTTTCGAAAGTTCTTAGCAAAATAAGCTAGTGCTT
CTGTCACAATATGACTACCAATCCCTTTCTCTGATAGGCTTGATCAACCATAAACAAC
CAATAAAAACAGTCTCCTCATCAGGATATGCATAGACAAAATCCATAACAGCCACAAGGT
CAAATCCATTCCAAAATCCAACAAAAAATTATCAGCCTTAGCTTTACCTTCAGGTAGAC
AAAGCATGTCTCTTTTACAGTTGCAAAATTTGGCTCTGGTGGACAATGCTGAAAATACA
GAGGATTACTTTCATATAAAGATAAAATACTTGAATATCCTTTTCAGTTAGTATCCTAC
AACTGTAATACTTAGATAGTTGGTCAATCATCTTTCAAATTCGATACTTTCTTGCGCC
TGTGATTATGACACAGGAAGATGCACTGATCGTCATCAGCCACATAAAAGTTCTTTCCAT
CGTGCCTAATCGTTGTCTCAAACCTTTGGATAAAACCTTTAGCCTATACAACCTGGATTTT
CCTCTCTCAAAGTATATTCTTTTGCAGGCGAACTTCCTCAAAATCAGTCGTGTGCAACT
TCAGTAGAATATTCATAGGCTCGGATAATCTGAGCGACAACAGGATGGCGAACCACATCC
TTGGCTGAAAAATGAACAAAGTCAATCTGATGGATGTTCTTGAGTTTCTCTTGAGCATCA

ATCAAACCGGACTTGACATTACGTGGCAGGTCAATCTGACTAATA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 431 | 1003 | R | 191 aa |

> 3864862-1 ORF translation from 431-1003, direction R
 VADDDQCIFLCHNHRAQESIEFEKIDQLSKYYSRILTEKDIPSILSLYESNPLYFQHC
 PPEPNFATVKEDMLCLPEGKAKADKFFVGFWNGFDLVAVMDVFVYAYPDEETVFIGLFMVD
 QAYQQRKGIGSHIVTEALAYFAKNFRKARLAYVKGNPQSQHFWEKQGFKSIGCEVKQELYT
 VVIVEQSLED*

Description:
 unknown

Assembly ID: 3864888
 Assembly Length: 1742bp

> 3864888 Strep Assembly -- Assembly id#3864888
 CTAATCTCCTTAAAACGTGATCTTTTCAAGAATATTTTATCTAAACAATCCAGCAAGTC
 TTGGTAAGAATAGACTTCGTAAGTCGGCTGGGCTTGTGTGTGATTTTCGAGGTGATGAGG
 ATTATACCAGATAGTGTCAATCCCCGCATTATTGCCACCTTGAATGTCGGCGGTTAGAGA
 ATCTCCAATCATCAGCGTCTTTTCTTTACTAAATCCAGCAATTTGCTGGCCAATCTTTTC
 ATAAAAAAGAGCATCCGGCTTTTGAGTTTGCAACTGTTCTGAGATAAAGACTTGATTGAA
 ATAAGGTGCTAGACCAGATTGAGCCAAACGTCCTGTCTGAATGGCAGTAATGCCATTTGT
 CGCAGCATACAAGTTATAATCACGCTCAATGAGGCTGTCCAAGAGATCATGAGCGCCCGA
 TAGTGTTTGTCCCTGCTGGGCGAGGTAATTTGGTAACGCTGGGCAAGAAAACCTACCGTC
 TTTTCTCTGTCCAAATGAGCAAATAAACGAGAAAAGCGCGTGTTAACCAGCTCTTGTTT
 ACTGATTTTCTTCAGCTCCAAGTCTTTCCAGAGAGCCTTGTTTCATAGGAACGTAATAATC
 TTTATAAGCCGGAATATCCGCAACTCCTTCTTTCTTTTAGAAGTGGAGTCAAAGCCACATC
 CTCAGCAGCATCAAAATCAAGAAGAGTGTGGTCGAGGTGGAAGAGTACAAATTTGTAGAA
 CAATTTGAGGTTTTCTTTCTGAAAATTCATTAAGAACATTATATCATAAAGCACCTCAT
 ACAATTAACATAATTTAATCACTTAAAAAAATTCGAACACTTTCTATACAACTGACAGCT
 CAAATCTTTTCAGAATAGAACAATACTAATACTATCGAACACCCCGTCTTCATAAATACATAT
 GTAATTCTAGGCCTAGAATTCCCTATAAACTAAATGCTTTTCATACTCTTCCAAGTAATTGA
 TTGCCTTAAATTTTAATTTTGAAGGTTTCTAAAGCTAGAATAGCCCCATCACAAATCAGT
 TTTGATTGATTACAAATTTAGAAACACTATAGTTTCACTCCTGTTAAAAATAAAAGGAAC
 TGCATAAAGCAATCCCTTTCTGATTTTGAAATCATTTACTTAACATTTTATAGTTGAGAT
 AATCAATAGCTTATCTATAAAAAGAGTTATAGTAAATTCCTTATTTATTGATTCCAAGC
 TCCGCTAACTGTATTTGAATAACTGACAGTTCTGCACCAGCCTGAAAAAGAGCAGCTGCA
 TTATAGGCACCTTCTACAATTGGAACCCTGTTGATGATGATACTTTTATCACTGAAATCA
 GTCACCATTTTTTAAGTTCATTTTAGCAGAACCCTAGGTCAAAAAGGCAAGTAAAGTATCT

GCTGGATTTTCGGAAACAACCCTATCTACTTGATCAAACTCGTTCCAATTCCCTCCGCC
 TCGGTTCCCTCCTACATAAGTAATCGGAACATCTTTAGCTACTTTACTAATCAGTTCAACA
 ACACCTTCTGCAATGTGTTTGAATGTGAAACGATAACAAGACCAATACCAATACTTTCC
 ATCAAACCACTCCAGTTTCTAAAATAGCAGTAAAGAGTAATCCTGATGAGAATGATCCAG
 GATCAATATGTCCAAGAAACCACATGCTCCTAAGACAAGAGCTAACAGACTGGCCATCAA
 TAATAGTATTGTTCTTTTTTTCATCATTACTCCTTAAGTAGTGTTTAACTGATTAATTTCG
 AT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 10 | 657 | R | 216 aa |

> 3864888-1 ORF translation from 10-657, direction R
 VALTPLLLKEEGVADIPAYKDYYVPMNKALWKDLELKKISKQELVNTRFSRLFAHFGQEKD
 GSFLAQRYQFYLAQQGQTLGSAHDLLDSLIERDYNLYAATNGITAIQTGRLAQSGGLAPYF
 NQVFISEQLQTQKPDALFYEKIGQQIAGFSKEKTLMIGDSLTDIQQGNNAGIDTIWYNP
 HHLENHTQAQPTYEVYSYQDLLDCLDKNILEKITF*

Description:

unknown

Assembly ID: 3864898

Assembly Length: 1136bp

> 3864898 Strep Assembly -- Assembly id#3864898
 GTGGAATGCGGGGACGCCTTGTCTAATTTTGGATCAAGCCCTGAGTTTGACACAGGGAAA
 TGAGCTGGACGGACTGCTATCTCTGAAGAAATTACTGGCACCATTAGCCTATCAGCCTTG
 GATGATTATGTGGCGGCCTTGTCTCAACAGGATGTTCCCAAAGCTTTGTCTTGCTTGAAT
 CTTCTTTTTGACAATGGTAAGAGCATGACTCGTTTTGTGACCGATCTTTTGCACATTTTA
 AGAGACTTGTTAATTGTTCAAACAGGGGGAGAAAATACTCATCATAGTTCAGTCTTTGTA
 GAAAATTTGGCACTTCCTCAAAAAAATCTGTTTGAAATGATTGCTTAGCAACAGTGAAT
 TTAGCAGATATTAAGTCTAGTTTGCAGCCCAAGATTTATGCTGAAATGATGACCGTCCGT
 TTGGCGGAAATCAAGCCCGAACCAGCTCTATCAGGAGCGGTTGAAAATCGAATTGCTACG
 CTGAGACAGGAAGTTGCCCGTCTCAAACAAGAGCTTCTAATGCAGGTGCGGTTCTTAA
 CAAGTTGCACCAGCTCCTAGTCGACCAGCTACGGGCAAACAGTCTATCGTGTCGATCGC
 AATAAAGTGCAATCTATCTTACAAGAGGCCGTCGAAAATCCTGATTTAGCACGTCAAAAT
 CTAATTCGTTTGCAGAATGCCTGGGGAGAGGTAATTGAAAGTCTAGGTGGGCCGGACAAG
 GCTCTGCTAGTTGGTTCTCAACCGGTTGCTGCCAATGAACACCATGCTATTCTTGCTTTT
 GAGTCTAACTTCAATGCTGGTCAAACATATGAAACGAGACAATCTCAATACCATGTTTGGT
 AATATCCTCAGTCAGGCGGCAGGTTTTTCACCTGAGATTTTAGCTATTTCCATGGAGGAA
 TGGAAAGAAGTTCGCGCAGCCTTTTCAGCCAAAGCCAAATCTTCTCAAACGAAAAAGAA
 GTAGAAGAAAGCCTGATTCAGAAGGATTTGAATTTTGGCTGATAAAGTGAAGGTAGAG

GAAGACTAAAGAAAGATTTTCATGATACAATAAGTTTATGAATAAACAACAATTTATTATTT
ATGGCGCTATTTACAGCTGCTGAGACCTATTTTTTCAATGAAGCCTGGATGACTGG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 130 | 1029 | F | 300 aa |

> 3864898-1 ORF translation from 130-1029, direction F
VAALSQQDVPKALSCLNLLFDNGKSMTRFVTDLLHYLRDLLIVQTGGENTHHSSVFVENL
ALPQKNLFEMIRLATVNLADIKSSLQPKIYAEMMTVRLAEIKPEPALSGAVENRIATLRQ
EVARLKQELSNAGAVPKQVAPAPSRPATGKTVYRVDRNKVQSILQEAVENPDLARQNLIR
LQNAWGEVIESLGGPDKALLVGSQPVAANEHHAILAFESNFNAGQTMKRDNLNTMFGNIL
SQAAGFSPEILAI SMEEWKEVRAAFSAKAKSSQTEKEVEESLIPEGFEFLADKVKVEED*

Description:

unknown

Assembly ID: 3864938

Assembly Length: 1670bp

> 3864938 Strep Assembly -- Assembly id#3864938

CTGTCTCTGAAACAGTCACATCAAGTGCCTCTGAACAANCGCCCCNCCTAGGTNGACGGT
ATCGATAAGCTCGATCTGTGATTTTCAGAGAAGAAATCAAGTGCTGTAACAGAAGTAAGAT
GTAATTGTATGTAAAGGAGACGTCATGTTAAATAGTATTGTAACCATTTATTGTATTGCC
CTTATCGCGTTTATCTTGTTTTGGTTTTTCAAAAAGCCTGAAAAATCTGGACAAAAAGCC
CAGCAAAAAAACGGATACCAAGAGATTCGAGTGGAAGTCATGGGAGGCTATACTCCTGAG
TTGATTGTCCCTCAAGAAATCAGTGCCAGCCCGCATTGTCTTTGACCGCAAGGATCCTTCA
CCATGTCTGGATCAAATTGTTTTTCCAGATTTTGGTGTACATGCGAACCTGCCAATGGGG
GAAGAGTATGTAGTGGAATCACGCCCTGAACAGGCTGGAGAGTTTGGCTTTGCTTGTGGT
ATGAACATGATGCACGGCAAGATGATTGTAGAGTAGGTGGAGACTATGACAGAAATTGTG
AAAGCAAGCTTAGAAAAATGGCATTCAAAAAATCCGTATCCGAGCTGAAAAAGGCTATCAT
CCAGCCCATATCCAGCTTCAAAGGGAATTCCAGCTGAGATTACCTTTTCATTTCGTGCTAC
TCCTTCAAACGTGTTATAAGGGAAATCTGTTTGAAGAAGAAGGTATCTTGGAAGCAATCG
GCGTAGATGAGGAGAAAGTCATTTCGTTTTACACCTCAAGAATTAGGGAGACATGAATTTT
CTTGTGGCATGAAGATGCAAAGGGAAGCTATATAGTCGTTGAGAAGACTCGAAAATCTC
TATCTCTCCTGCAAACGTTTTTGGATTACTAGTATCTTTACTGTGCCTCTTGTGATTCTC
ATGATTGGGATGTTGGCAGGTAGCATTAGTCATCAAGTCATGCATTGGGGAACCTTTTTA
GCAACAACGCCTATTATGTTAGTTGCGGGTAAGCCATATATCCAGAGTGCTTGGGCCAGT
TTTAAAAAGCACAATGCCAACATGGATACCTTGGTTGCGCTGGGAACTCTAGTGGCTTAT
TTCTATAGCCTAGTTGCTCTCTTTGCTGGTCTCCCTGTTTACTTCGAAAGTGCTGGATTT
ATCCTCTTTTTCGTTCTTTTGGGAGCAGTTTTTGAGGAAAAAATGAGGAAAAAATACGTCC
CAAGCTGTGGAGAAATTACTGGACTTGCAAGCTAAAACCGCAGAAGTCTTGAGTGATGAT

AGTTATGTCCAAGTTCCTTTGGAACAAGTCAAGGTACGCGACCTTGATTCCAGTGCGTCC
 CGGTGAAAAGATTGCTGTTGATGGTGTCTAGTAGAAGGTGTCTCTAGTATTGACGAATC
 CATGGTGACAGGTGAGAGTCTGCCTGTGGACAAGACAGTTGGAGATACTGTCATTGGCTC
 AACCATCAATCATAGTGGAACGCTTGTCTTTAGAGCAGAAAAAGTTGGCTCAGAGACTGT
 TTTGGCTCAGATTGTAGATTTTGTGAAGAAAGCTCAGACAAGTCGTGCGCCGATTCAGGA
 CTTGACGGATAAGATTTTCAGGGATTTTGTCCCAGTAGTTGTCATTTTAGGAATCATGAC
 CTTTTGGGTTTGGTTTCGTCTTGCTCAGGGATAGTGTGGTCGTGCTTGGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 883 | 1326 | F | 148 aa |

> 3864938-2 ORF translation from 883-1326, direction F
 VPLVILMIGMLAGSISHQVMHWGTFLATTPIMLVAGKPYIQSAWASFKKHNANMDTLVAL
 GTLVAYFYSLVALFAGLPVYFESAGFILFFVLLGAVFEEMRKNTSQAVEKLLDLQAKTA
 EVLSDDSYVQVPLEQVKVRDLDSASR*

Description:

ATCS_SYNP7

Assembly ID: 3864956

Assembly Length: 1252bp

> 3864956 Strep Assembly -- Assembly id#3864956

ACAAGAACAATTGGAACAGGTACAGGCTGTTAAAAATCGATTAAACACAGCTAGTGAAGA
 AGTGAAAAACCAAGTCTTGCTACCCATGGCTGATCACTTAGTGGCTGCTACTGAGGAAAT
 TTTAGCGGCTAATGCCCTCGATATGGCAGCGGCTAAGGGGAAAATCTCAGATGTGATGTT
 GGATCGTCTTTATTTGGATGCAGATCGTATAGAAGCGATGGCAAGAGGAATTCGTGAAGT
 GGTTGCCCTTACCAGATCCAATCGGTGAAGTTTTAGAAACAAGTCAGCTTGAAAATGGTTT
 GGTTATCACAAAAAACGTGTAGCTATGGGGGTCATCGGTATTATCTATGAAAGCCGTCC
 AAATGTGACGTCTGATGCGGCTGCTTTGACTCTTAAGAGTGGAATGCGGTTGTTCTTCG
 TAGTGGTAAGGATGCCTATCAAACAACCCATGCCATTGTACAGCCTTGAAGAAGGGCTT
 GGAGACGACTACTATTCATCCAAATGTGATTCAACTGGTGGAGGATACTAGCCGTGAAAG
 TAGTTATGCTATGATGAAGGCCAAGGGCTATCTAGACCTTCTCATTCCTCGTGGAGGAGC
 TGGCTTGATTAATGCAGTAGTTGAGAATGCCATTGTGCCTGTTATCGAGACAGGAAGTGG
 GATTGTCCATGTTTATGTGCGATAAGGACGCAGATGACGACAAGGCACTGTCTATCATCAA
 CAATGCCAAAACCAAGTCGTCCTTCTGTCTGCAATGCCATGGAGGTTCTGCTGGTTCATGA
 AGACAAGGCAGCAAGCTTCCTTCCTCGCTTGGAGCAAGTGCTGGTTGCAGATCGAAAAGA
 AGCTGGGTTGGAACCAATTCAATTCCGCCTAGATAGCAAAGCAAGCCAGTTTGTTCAGG
 TCAAGCTGCTCAAGCACAGACTTTGATACCGAGTTTTTAGACTATATTCTAGCTGTAA
 GGTTGTGAGCAGTTTAGAAGAAGCGGTTGCGCATATTGAATCCACAGTACCCATCATTCG
 GATGCTATTGTGACGGAAAATGCTGAAGCTGCAGCATACTTTACAGATCAAGTGGACTCT

GCAGCGGTGTATGTTAATGCCTCAACTCGTTTCACAGATGGAGGACAATTTGGTCTTGGT
 TGTGAAATGGGGATTTCTACTCAGAAATTGCACGCGCGTGGTCCAATGGGCTTGAAAGAG
 TTGACCAGCTACAAGTATGTGGTTGCTGGTGATGGGCAGATAAGGGAGTAAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1030 | 1251 | F | 74 aa |

> 3864956-2 ORF translation from 1030-1251, direction F
 VTENAEAAAYFTDQVDSAAVYVNASTRFTDGGQFGLGCEMGISTQKLHARGPMGLKELTS
 YKYVVAGDGQIRE*

Description:

gamma-glutamyl phosphate reductase (proA) homolog - Haemophilus
 influenzae (str ain Rd KW20)

Assembly ID: 3864958

Assembly Length: 1785bp

> 3864958 Strep Assembly -- Assembly id#3864958

CTGCCCTAGCAGGAACGCAAGAAGGAAGTGGAGAATAGGCATTTTCAAATTATAACCTA
 CACTAGCCATCATATCTAATGTTGGAGTGCTAACTAGCTTATCCTTACTATTCAAGGATA
 AGGCGTCTGCTCTCATTTGATCTACAACAATCAAATAATATTTGGTTGTTTTGTCTGAA
 CCATAAAATCTCCTTTCTAATATGGCAAAGAGGCACAAGAAGATATCTACCTTTACTGC
 ACCCCTTTCTATATCAATCTCTCTATATAAAGCAATAACATTCTTGTATGTTTTATAGA
 ACAATGGACTAAAATATGACTAAATCGATTAGGAAATTCAAATCATTTTCTAGTACTGTT
 TTAGTAAGTTACAGTGTAATCTTCCAACCTCAATAAATTATAAACCTTTGTCTAATAACA
 ATTTTAGTGGAGATAAGAAATCCTACACCTAACTCATCTTACACGTAATCTATTTCTATT
 TTATCACAAAAACGCAAGTAAGACCATTAACTCAATTCAGTTTTATCTGCCATTTTCAC
 AAATGGGAAATAAGTCAAGACACTAATAATCAAACAAACAAGTATAAGATGATGGCAGC
 CCAATCAAATGCTGTAGAGAAGAAACCATATAAAATTGGAGGCATTACCCAAGTAACATT
 TTGTGTAACAGGTGAAACAAGACCCAGCTTGTTGCCAGTAAGCTACCGTTGCCATGAA
 AACCGGGCTAAGTACAAATGGTATAAATAGCAAAGGATTCAAGACAAGTGGTAAACCATA
 ATTCGATACCGGCTCACCAATATTAAACAGAACTGGTGCTAGACCAAGTTTAGCAACTTT
 TCGATAATGACTGTTTCTTGAAAAAATTAAAATAGCAAGTACTAATCCTAATCCTCCAAA
 CCAGACAAACGCCCCAAAAGACCCACTTGTCATATATAAGGAATCGGTTACCTTTTTG
 GAAAGCATCCAGATTCGCTAACATAGCAACTCCAAATAGCCCTTCCATGATGGGAGCCAA
 TACATTTCTCCATGGAGACCAAAAAACCAGAATAACTTATTCAAAAAGATCATCAGAAT
 AACTGCAAAGAACTTTGAGACAAACCTAGTAATGGCGTTTGTAACACCTTGTAACCCCA
 ATCAATCAATAAGTCATTGCTAAGTAAATGGAAAACATAAGTCAAGATGGCTACTATATA
 CATCGCCATAAATCCTGGAATGATAGAAGTGAACGGCTTAGCAATCGCAGGGGGAAGTGA
 ATCTGGTAACTTGATTACCCAGTTCTTTTTCATTACTTTACAGAAAATAATAGAGGCTAA

AAATCCAATCATCATGGCTGTAAAGTAGCCTCTGGCATTAAATATGGTTTCCTGGAATCAC
 ATTCCCAATAGTTACCATCAGATTTTTTACCATCAAATGCTAGATTATCAATTCCATGTTA
 AGATTTGATCTAATTTACATCTCCTACATTTGCCAAAGGGAAACTCTTTGTAAGTGTAC
 TTCCAATCGAAATGACAAACGAAGCAAGTGATACCAAACCAGCAGAAACTGTATCAACCT
 TGTAAATCTTAGCGATATTCACCTCCAAGCAATAGATGAACAACAAGGAAACAATTGGTA
 TACTTCCCTTGAATACCAAATTATTGATGTCAACAAGCCACTGAAAGGTTTTTCGTAATAC
 TTCTAGGTGAAATTGTTGTGGTAAATCCACTAGAAAAGCATTTAATAACAAAGCAATGG
 AACCTGTCATAATAACAGGCATAGTCCCCACAAATGAATCACGTT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 1427 | 1711 | R | 95 aa |

> 3864958-2 ORF translation from 1427-1711, direction R
 VDLPPQQFHLGSITKTFQWLVDINNLFVKGSIPVSLFLFIYCLGVNIAKIYKVDTVSAGLV
 SLASFVISIGSTVTKSFPLANVGDKLDQILTWN*

Description:

unknown

Assembly ID: 3865022

Assembly Length: 1386bp

> 3865022 Strep Assembly -- Assembly id#3865022

ATCGAATTTTCATTTCTATTTTCCTATTCCATTTTTTATTCAAAAAATCAAAAAGCAAACCTAG
 AAAGCTGGTCGCTGGTGGTTCAAACACTGTTTTGAGATTGTCAATAGAAGTACAAACC
 CTGTAATATACCTGCATATATACATACGACAAGGCGATACTACCCTAGTTTGAAGAGATT
 TTCGAAGAGTATTCATTTTTGTCTTTTACTTATTATACCATATTCACATAAAAAACGAA
 CATTTCTTATCCTAAAAATGCTCATTTTTCTTAAATTATCAATCTAAATCTGGTTTATAG
 AAGGAACGATTATCCATAGCGAAGATTTTATTGGTCATCTCTCCTTTATCCACCAAAGCC
 AGAGCTGTTGACATCATCATCATGCTTGCATCCAGATTGTCAATCATATGGATAATCTCT
 GCCTCCATAATACGTGGACGGACTGGAATTTCCATATTCAGCAAGCCGTGGTGGACTTG
 AGGATGACATGACGAAGCAAAACGACTTCTTCCTTGGTATCATCGATGCCGAGTTCCATA
 ACTGTCTTGGTAATTTGCTATCAATGAGAGCGATATGTCCAAGAAGATTACCTCGCACT
 GTGTACTCTGTCTGGTCTGGCCCCGTCAACTCGATAACCTTAGCTAAGTCATGCAGCATA
 ATCCCCGCATAGAGCAGGCTCTTATTGAGCTGAGGATAAACTTCGCTAATAGCGTCTGCC
 AAACGTACCATGGTCGCGGTATGATAAGCCAACCCCGTTTCAAAGGCATGGTGGTGGTC
 TTGGCGGCTGGATAGGAGTAGAATTCCTTATCATACTTGGTGTAGAGATTTTCGGACAATC
 CGTTGCCAGACAGGATTTTCAATTTTGAAAATCATTGCGACATGTAGTCACGAATTTCC
 TTGACATCAACTGGTGACTTGACCTTGAAATCAGCTGGGTGATTGGGTTTACCAGCTTGA
 GGCAGGCGGAGAGTAATTTGATTGACTTGAGGGGTATTGTTATAAACTTCTCGGCGTCCCT
 TTCATGTGGACAACCTTACCTGCGGTAAAGGCCTCAATGTTATGAGGTTGGGCATCCCAG

AGCTTCCCATCAATCTCGCCACTATCATCTTGGAAGGTAAAGGCTAGGTAGTTTTTCCCA
 GCTCGAGTTTGCCTCAGGTCAGCTGATTTGATTAGGTAAAAGCCTTCAAATAACTCATCT
 TTTTTCATGTGACTAATCTTCATATTCTTCCTCATTTTCTTGAAAATGGAGTAGATCAAG
 CGCAGGCTCACCTTCTGACAACTCAATGTGACGGAGCGTCCGCTCGATAGCTATGGTACG
 ACGGTTTAATAATTCGATCAATATTGCCAGAGGCATGTTGGAGATGTTTTGTGCCTTGA
 CCAGAA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 279 | 1271 | R | 331 aa |

> 3865022-1 ORF translation from 279-1271, direction R
 VSLRLIYSIFKKMRKNMKISHMKKDELFEFGFYLIKSAADLRQTRAGKNYLAFTFQDDSGEI
 DGKLWDAQPHNIEAFTAGKVVMKGRREVYNNTPOVNQITLRLPQAGEPNDPADFKVKSP
 VDVKEIRDYMSQMIFKIENPVWQRIVRNLYTKYDKEFYSPAAKTNHHAFETGLAYHTAT
 MURLADAISEVYPQLNKSLLYAGIMLHDLAKVIELTGPDPQTEYTVRGNLLGHIALIDSEI
 TKTMELGIDDTKEEVLLRHVILKSTTACLNMEIPVRPRIMEAEIHMIDNLDASMMM
 STALALVDKGEMTNKIFAMDNRSFYKPDLD*

Description:

gi|710422 (U21636) cmp-binding-factor 1 [Staphylococcus aureus]

Assembly ID: 3865036

Assembly Length: 1167bp

> 3865036 Strep Assembly -- Assembly id#3865036

CTCAGATTACAGAGGACAATCAACTGGTTCATTTTCGTTTCCAGTTTCAAAAAGGCTTAG
 AAAGGGAGTTCATCTATCGTGTGGAAAAAGAAAAAGTTAAGGCAGGTGTTCTCCTCTAC
 GCAGTCACCATAGCAGCCATCTTTAGTCTTTTGTTGCAATTTTATTTGAACCGACAAGTC
 GCCCACTATCAAGACTATGCTTTGAATAAAGAAAAATTGGTTGCTTTTGCTATGGCTAAA
 CGAACCAGATAAGGTTGAGCAAGAAAGTGGGGAACAGGTTTTTAATCTAGGTCAGGTA
 AGCTATCAAAACAAGAAAACTGGCTTAGTGACGAGGGTTCGTACGGATAAGAGCCAATAT
 GAGTTTCTGTTTCCTTCAGTCAAATCAAAGAAGAGAAAAGAGATAAAAAGGAAGAGGTA
 GCGACCGATTCAAGCGAAAAAGTGGAGAAGAAAAATCAGAAGAGAAGCCTGAAAAGAAA
 GAGAATTCCCTAGTCAATTCAACTATAATGCGTTGAATCCAGAATAGTCCACTGTAGTTTC
 TAGAAAATTGCTGGAAATGGATGTTAAGCTCCAATTCATTTGTTTATATCTTATTTTCAGT
 CCACTATACTTTGTGCTAAATTAAAGATATGAAACATGATTTTAACCACAAAGCAGAAAC
 TTTCGATTTCCCTAAAAATATCTTCCTCGCAAACCTGGTATGTCAAGCAGCCGAGAAACA
 GATTGATCTTCTATCAGACAAAGAAATTTTAGATTTCGGTGGTGGCAGGGTCTATTAGC
 CTTGCCCTTAACCCCTAGCCAAGCAGGCTAAGTCAGTCACTCTTGTAGACATTTCTGAGA
 AAATGTTGGAGCAAGCTCGTTTGAAAGTGGAGCAGCAAGCAATCAAGAATATCCAGTTTT
 TGGAGCAAGATTTACCGAAAAATCCCTTGGAGAAAGAGTTTGATTGCCTTGCTGTTAGTC

GGGTTCTTCATCATATGCCTGATTTGGATGCGGCTCTCTCACTGTTTCATCAACATTTGA
 AGGAAGATGGGAAACTCATCATTGCTGATTTTACCAAGACAGAAGCTAATCATCATGGAT
 TTGATTTAGCTGAACTGGAAAACAAGCTAATTGAGCATGGGTTTTTCATCTGTGCATAGT
 CAGATNCTCTATAGCGCTGAAGANCTG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 79 | 492 | F | 138 aa |

> 3865036-1 ORF translation from 79-492, direction F
 VWKKKKVKAGVLLYAVTIAAIFSLLLQFYLNQVAHYQDYALNKEKLVAFAMAKRTKDKV
 EQESGEQVFNLGQVSYQNKKTGLVTRVRTDKSQYEFLFPSVKIKEEKRDKKEEVATDSSE
 KVEKKKSEEKPEKKENS*

Description:

unknown

Assembly ID: 3865054

Assembly Length: 916bp

> 3865054 Strep Assembly -- Assembly id#3865054

TCTCCCAACATATAATTTCCGTTTTTCCAATCCCCCAGCTGTCATACAGTCTGTGATAAGA
 GCGATGTTTTCTGTTCTTTTGTGTTGATAAGAATTTGCAAGCCTTTGGATCTACGTGG
 TGACCATCACAGATCAACTCTGCATAGGTATGTGGCAATTGGTACATGGCTCCAACCATA
 CCAATTTCACGGTGAGTCAACCCACGCATTCCATTGTAGGCATGCACCCAAACACTCGCT
 CCAGCATCGACTGCTTTTTTGGCTTCATCAAAAGTCGCGTTTGAATGTCCAAGAGCAACC
 GTCACACCTTCGCCCCGTAAGTGTACGAACAAAGTCTTCCACCCCATCACGTTCTGGTGCA
 ATCGAATTTTATTAAGCAAGCCATTTGCCGCTTTTGGCCAAGAATGAAACTCCTCAACAC
 CCGGGTCTCTCATATAAGTTGGATTTTGTGCCCCCTTAAAAGTTTCTGTGAAATATGGAC
 CTTTATAATAAATCCCACGAATCTTAGCACCTGTTGCTTCTTTATAATGGTTTCCAAGAT
 TTTCAGTGACTGCAAGCAATTGCTCATAAGTGGCTGTTAAAGTTGTGGGTAAGAACTGG
 TAACACCGGTACTAAGAAGTCCCTCACTCATAGTATGCAATGTACCTTCAATGTTGTTGT
 CCATCACATCTACACCTGCATATCCATGAATATGAGTATCCACAAGACCTGGGGCAATGC
 TATAACCTGTATAGTCAATCACCTCAGCCCCCTCAGGAATCTGCTCTACATGTTTCCCAA
 ACTTGCCGTCCACAAGTTCCAAGTAACCACCTCGACAAATCCGTGTGGGTAGAAAACTG
 ATCCGCTTTAATATAGTTAGGCATAATGTTAACCTCCTTAAAAGATTGATTCTACAATTT
 ATTATGTCAATTCGAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 302 | 793 | R | 164 aa |

> 3865054-1 ORF translation from 302-793, direction R
 VDGKFGKHVEQIPEGAEVIDYTGYSIAPGLVDTHIHGYAGVDVMDNNIEGTLHTMSEGLL
 STGVTSTFLPTTLTATYEQLLAVTENLGNHYKEATGAKIRGIYYEGPYFTETFKGAQNPTY
 MRDPGVVEEFHSWQKAANGLLNKIRLHQNVMGWKTFLVQLRAKV*

Description:

N-acetylglucosamine-6-phosphate deacetylase (nagA) homolog -
 Haemophilus influe nzae (strain Rd KW20)

Assembly ID: 3865102

Assembly Length: 786bp

> 3865102 Strep Assembly -- Assembly id#3865102
 CTGGATTAAAACGAGGCAGTTTCAGACTAATATCCAAGTCGTAAGAAATGCCTGAAATAA
 GCTTTTCTAAATTGTCCAAAGCTTGCGGGAAAACGCTCTTGGAATAGTTTCTCTAAAGAA
 CTTGCTGATATAAAGACATCTTGCTCTCGAACGCAAGGGAAGTCTCTGAGCGGTAGATTT
 TCTTTAATCGCTGTTAAACTTGAAGAACTTCTCTATCCCTGCTTTCAAAAGCGTTGACC
 CGATAAAGAGGTAAGATAGGATGATGAAATTCGCTTGCTAGTGTTTCTGGATAAACCCCT
 ATATAGTAATCACAGCCTAGTTCTAACGACTCAACTCTATCAAATAAGGCACAATGACC
 GCGATATCCTCCAGGTACTGGGACAGGACTGACCAAGTTTTCTCCCCCTGCATCTTGCGT
 GTCGAAAGCTTCATCAACTGCTGATAGCCACACTAGATAGAGCTAAAAAGCGCAAATTC
 ACTTCCTGATCATCTACAAACACTGTCATTTCAAGCCCTAGCAAAGGATGAATGCCGTAT
 TTTTTTGTAATCTCTAGAAAGTCGAAAGCGCCATAAAGATTGTCAATATCCATCATAGCC
 AAATGAGTGTAGCCGTATTCTTTAGCTGCTCTCACATACTTTTCGATCGAAATGACGCTT
 TCCATAAACTATAGACTGTTTTTGTATCTAGTTGTGCGATCAATTTACACTTCTCCTCT
 ATCCTTCTCACTATATTATACCATTTTCACCTATAAATGGCTTCTCTTGAGAAAAATTTC
 GATCAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 27 | 731 | R | 235 aa |

> 3865102-1 ORF translation from 27-231, direction R
 VRRIEEKCKLIAQLDTKTVYSFMESVISIEKYVRAAKEYGYTHLAMMDIDNLYGAFDFLE
 ITKKYGIHPLLGLEMTVFVDDQEVNLRFLALSSVGYYQLMKLSTAKMQGEKTWSVLSQYL
 EDIAVIVPYFDRVESLELGCDYYIGVYPETLASEFHHPILPLYRVNAFESRDREVLQVLT
 AIKENLPLREVPLRSRQDVFISSSLEKLFQERFPASFGQFRKAYFRHFLRLGY*

Description:

unknown

Assembly ID: 3865156

Assembly Length: 1213bp

```
> 3865156 Strep Assembly -- Assembly id#3865156
CACTTTCAGCTTCTTCTCTTTTGAACGGTTATAAACACGAATCAGATTCCCTATTTCTT
GCGATTTATGTGATTTCCTTATTTTCCAATCTAAAGTATAGTGAAATGAAATAAACATGC
GCAAATCGATTAAGGAATTTAATCTAATTTCTAACAATGTCTTAGAAATCAAAGTGACT
ATTTTAACCTCAATGCACTAAACATCTAATACTCAATAAAAATCAAAGAGCAAAGTAGGA
AACTAGCCGCAGGTGGCTCAAAACACTGTTTTGAGGTTGTAGATGAAACTGACGAAGTCA
GTAACCATACATACGGCAAGGCGACGCTGACGTGGTTTGAAGAGATTTTCGAAGAGTAGC
AAAATGGAAAAAGGAGTGAGTGAAGCACATCGCCTCCCCACTCCTTTTTCTGTTTTTAGG
CTGTTTTTTCAACCTTCAAGATTTTTTACATCATAGCTACCAACAGGCGTTTCAATGGTTG
CTGTATCACCTGTTTTCTTGCCAATCAAGGCCTGCCCAATTGGGCTTTCATTTGAAACCT
TACCTGCAAAGGCATCCGCACCAGCTGAACCTACGATAATATAAACTTCTTCTTCGTCCT
CACCAATTTCTTGGATGGTGACTGTTTTACCAATCGCTACTTCGTCCTGGGCAACTGCGT
CGCTATTGACGATTTTCAAGCATAGCGGATTTTTGTTTTCTAAGCTAGAGATTTGTCCTTCGA
CAAAGGCTTGTTTCATCCTTAGCTGCTTCGTACTCACTGTTTTCTGAAAGGTCACCGTATG
AACGGGCAATCTTAATGCGTTCTACCACTTCTGGTTCGACGAAACCAATTTCAATTCTTCT
AATTCTTTTTCAAGTTTTTTCCTTTTCTCAAGGGTCATAGGATATGTTTTTTCTGCCATT
TTTCTCAACTTTCTTCTGATAATATTTTCTAAAGAAAATTATGTGAAGTATCACATAATT
TTAGTTTGTTTAGTTTAATTTGCTGTTGACATGTTTCAAGCGACATTGCGGTCTGTTCTC
TTGATTGTTAGCATAGTAAACCTTGCTTCTGTGACATCTGCTACAAAGTAAAGTTATC
GCTCTTAGTTTGATTGATGCTTGACTCAATCCGCATCCAAGACTTGGACTATCGACTGGA
CCAGGCATGAGACCTACATTTTATAAACATTATAAGGTGAATCAATGTTGGTATCAATC
GCAACATCCTCAG
```

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 416 | 808 | R | 131 aa |

```
> 3865156-1 ORF translation from 416-808, direction R
VVERIKIARSYGDLSENSEYEAAKDEQAFVEGQISSLETKIRYAEIVNSDAVAQDEVAIG
KTVTIQEIGEDEEEVYIIIVGSAGADAFAGKVS NESPIGQALIGKKTGDTATIETPVGSYD
VKILKVEKTA*
```

Description:

TRANSCRIPTION ELONGATION FACTOR GRE (TRANSCRIPT CLEAVAGE FACTOR GRE). - ESCHERICHIA COLI.

Assembly ID: 3865160
Assembly Length: 1173bp

> 3865160 Strep Assembly -- Assembly id#3865160

TGCGGCTGAGTTGGGAATTCCTATCGTTAATAAGCGTGTATCGGTGACACCTATTTCTCT
 GATTGGGGCAGCGACAGATGCGACGGACTACTGGTTCTGGCAAAAGCGCTTGATAAGGCT
 GCGAAAGAGATTGGTGTGGACTTTATTGGTGGTCTTTCTGCCTTAGAACAAAAAGGTTAT
 CAAAAGGGAGATGAGATTCTCATCAATTCCATTCTCGCGCTTTGACTGAGACGGATAAG
 GTCTGCTCGTCAGTCAATATCGGCTCAACCAAGTCTGGTATTAATATGACGGCTGTGGCA
 GATATGGGACGAATTTATCAAGGAAACGGCAAATCTTTCAGATATGGGAGCGGCCAAGTT
 GGTGTATTTCGCTAATGCTGTTGAGGACAATCCATTTATGGCGGGTGCCTTTCATGGTGT
 TGGGGAAGCAGATGTTATCATCAATGTCGGAGTTTCTGGTCCTGGTGTGGTGAAACGTGC
 TTTGGAAAAAGTTCGTGGACAGAGCTTTGATGTTAGTAACCCGAAAACAGTTAAGAAAA
 CTGCCTTTTAAAATCACTCCGTATCCGGTCCAATTGGTTTGGTCAAATGCCCAGTGAGAG
 ACTGGGTGTGGAGTTTGGTATTGTGGACTTGAGTTTGGCACCAACCCCTGCGGTTGGAGA
 CTCTGTGGCACGTGTCCTTGAGGAAATGGGGCTAGAAACAGTTGGCACGCATGGAACGAC
 AGCTGCCTTGGCCCTCTTGAACGACCAAGTTAAAAAGGGTGGAGTGATGGCCTGTAACCA
 GGTGCGGTGGTCTATCTGGTGCCTTTATCCCTGTTTCTGAGGATGAAGGAATGATTGCTGC
 AGTGCAAAATGGCTCTCTTAATTTAGAAAACTAGAAGCTATGACGGCTATCTGTTCTTG
 TTGGATTGGATATGATTGCCATCCCGAAGATACGCCTGCTGAAACTATTGCGGCTATGA
 TTGCGGATGAAGCAGCAATCGGTGTTATCAACATGAAAACAACAGCTGTTTCGTATCATTC
 CCAAAGGAAGAGAAGGCGATATGATTGAGTTTGGTGGTCTATTAGGAAGTGCACCCGTTA
 TGAAGGTTAATGGGGCTTCGTCTGTCGACTTCATCTCTCGCGGTGGACAAATCCCAGCAC
 CAATTCATAGTTTTTAAAAATTAAGAAAATAGGA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 136 | 375 | F | 80 aa |

> 3865160-1 ORF translation from 136-375, direction F
 VDFIGGLSALEQKGYQKGDEILINSIPRALTETDKVCSSVNIGSTKSGINMTAVADMGR
 YQNGKSFYRYSQVGCIR*

Description:

unknown

Assembly ID: 3865172

Assembly Length: 1209bp

> 3865172 Strep Assembly -- Assembly id#3865172

TCGGAATCTGAGCTAGTGTAGCTTCCTTAATCTTATCTGATAAGATAGCTGTCATATCAG
 ACTCAATCATTTTCCTGGAGCAATCAACATTGACTCGTATATTCCGACTAGCGACCTCGCG
 TGCCACAGACTTGGTAAAGCCAATCAAGCCAGCCTTAGAAGCAGCATAGTTAGCTTGACC
 AATATTCCCCATCAAACCAACAACACTAGACATATTAATGATAGCACCTTCTCTGGCTTT
 CATCATCGGTTTCAAGACTGATTGTGTCATATTAAAGGCACCAGTCAGATTGACCTTGAG
 CACTTTTTTCAAATCTGCTTCTGTATCTTGAGCATAAGAGTATCTTGGGTAATCCCTGC

ATTGTTGACCAAAACATCTACTGAACCCAGTTCTGCAATAGCTTGATCAATCATACGCTT
 AGCGTCTGCAAAATCTGATACATCTCCTGAAATGGGAACCACCTTGATACCATAGTTTGA
 AAATCAGCGAGCAATTCTTCTGAGATTGCCCCACGACTGTTTAAGACAATGTTGGCTCC
 TGCTTGAGCAAACTTGTGGGCGATGGCAAGACCAATTCCACGACTCGAACCTGTAATAAA
 GATATTTTTATGTTCTAGTTTCATTTTTTTCCTTTCAAACTTCTACTTATTTTAGTCTA
 TTTTCTAAAAGTGCTACTAACTCGCTTGATCTTCCACATGAGCTAAGTGAGCAGTTTG
 ATCAATTTTTTTAACAAAACCTGACAAGACTTTCCCCGGTCCAATCTCGAATAAAGTTGC
 TTATGCCTGCTTCTTGCATGACCCCAATACTTTTCATAGAAACGAACGGGTTCCTTGACCT
 GACGCGTCAAGAGCTGAGCAATGTCCTCTTTTGCATCACAGCAGCTTCTGTATTGCCGA
 CTAGGGGACAAGTAAAATCTGAAAACTTACCTGAGCTAGAGTTTCAGCTAGTTTCTGGC
 TAGCAGGCTCAAGGAGAGCGGTGTGAAAGGGACCTGACACCTTAAGAGGAATCAAGCGTT
 TGGCACCTGCTTCTTGCAAAAGTTCAACCGCTCGATCAACTGCAACCACTTCTCCAGCAA
 TGACGATTTGTGCAGGTGTGTTATAGTTGGCTGGAGTAACCACTCCAAGTTCCAGAAGCT
 TTTTGACAGGCTTCTTCAATGACCTCTACTGGCGTATTGAGAACTGCTACCATCTTGCCA
 AGTTCAGCA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|------|-----------|--------|
| 1 | 731 | 1123 | R | 131 aa |

> 3865172-2 ORF translation from 731-1123, direction R

VVTPANYNTPAQIVIAGEVVAVDRAVELLQEAGAKRLIPLKVSGPFHTALLEPASQKLAE
 TLAQVSFSDFTCPLVGNTAAVMQKEDIAQLLTRQVKEPVRFYESIGVMQEAGISNFIRD
 WTGESLVRFC*

Description:

malonyl coenzyme A-acyl carrier protein transacylase (fabD)
 homolog - Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865228

Assembly Length: 813bp

> 3865228 Strep Assembly -- Assembly id#3865228

ATGACACGTCTGTTCTCTCAAGCAGAAATGGCAGAGTAACAAGCTCGATATTGAGGTAGC
 CGATAAAGAATTGGCTGAATTTGAAGCTCAGATTAAACAGGAAGTGGAAGCTCCAACCTTG
 TAGTGAGTCCTCAGGTTGAAGAAGAGCCTCAGCTCATCCAGTTGGCCCAATGTATGAAGA
 ACCAGAAGTAAATCCAGTGCATCCGACAGGTCCAACACCAGCTACAGAACTGTTGATTC
 AATACCGGGATTGTAAGCACCGCAAGAATCTGTTACAATTTTATAAGAAATATTCTGAGA
 ACAATATCTTATCCTTATATTTCCAGCGAGCAGGAAATGGTGTGAGTCCTGCATTCCCTA
 TCGATAAGATTATCCTCTCAAACTATCAAGTCTGAATCTAGTAAGATTTGACGTTCCCCA
 CGTTACGGGATAAGAGAGAGAAAGACTAAATCTTTTTCCGAATAAAGGTGGTACCACGAT
 TTTCGTCCTTTTTGGAAGTCGTGGTTTTTAATTTGTTATTATTTATAAAGGAGATACCAT

GAAACTCAAAGACACCCTTAATCTTGGGAAAACCTGAATTCCCAATGCGTGCAGGCCTTCC
 TACCAAAGAGCCAGTTTGGCAAAGGAATGGGAAGATGCAAACCTTTATCAACGTCGTCA
 AGAATTGAACCAAGGAAAACCTCATTTACCTTGCATGATGGCCCTCCATACGCTAACGG
 AAATATCCACGTTGGACATGCTATGAACAAGATTTCAAAGATATCATTTGTTTCGTTCTAA
 GTCTATGTCAGGATTTTACGCGCCATTTATTC

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 197 | 286 | F | 30 aa |

> 3865228-1 ORF translation from 197-286, direction F
 VHPTGPTPATETVDSIPGFEPQESVTIL*

Description:

unknown

Assembly ID: 3865230

Assembly Length: 953bp

> 3865230 Strep Assembly -- Assembly id#3865230

ATCGAATTATTTTGAACAAGGTGGATCAGCTATTTTGGCCTTGATTAGTATTTTACTCT
 TTAAATACACTTGAAGGTCGATTCTAATCTCGCTAATCCTTTTAAATCCAGAATAAGGGA
 AATATGTTATACTTGTTTTAAAGAAAAAGTTTCATTGAATTGGTTTTGAGGAGTTAGAA
 ATGAAAGTATTAGTGACAGGTTTTGAGCCCTTTGAGGCCATTAAAGGTTTACCAGCTGA
 AATCCATGGTGCTGAGGTCCGTTGGCTAGAGGTGCCGACAGTTTTTTCACAAATCTGCTCA
 AGTATTGGAAGAAGAGATGAATCGTTATCAACCTGACTTTGTCTTTGTATTGGGCAAGC
 TGGTGGGAAGAACTAGTTTGACACCTGAACGAGTGGCCATTAATCAAGACGATGCACGTAC
 TTCTGATAACGAAGATAATCAACCGATTGACCGTCCCATTCGCCCAGATGGTGCTTCGGC
 CTACTTTAGTAGTTTGCCGATTAAAGCGATGGTTCAAGCTATAAAAAAGAAGGATTACCG
 GCCTCTGTTTCCAATACGGCAGGGACTTTTGTCTGCAGCCATTTGATGTATCAGGCTCTC
 TATTTGGTAGAAAAGAAATTTCCCATATGTTAAGGCAGGTTTTATGCATATTCCTTATATG
 ATGGAACAGGTGGTGAACAGACCGACTACTCCAATATGAGTTTAGTGGATATTCGGCGA
 GGGATAGAAGCAGCAATCGGCGCTATGATAGAACATGGAGATCAGGAACCTCAAGTTGGTA
 GGCGGAGAAATTCATTGATAGAAAAAGCTTGAGGGGAAAACCTTCAAGCTTTTGGACGT
 TTTTCGAGCCAATACTGCTCGGTAAACATAATTTTAGTGCATTGGATATAAGGTAGGAGT
 GAAAAACTAGCAATGCCAAAGGTAATCCAATTGAGGAAGTACCAAGGAAGAAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 272 | 586 | F | 105 aa |

> 3865230-1 ORF translation from 272-586, direction F
 VPTVFHKS AQVLEEEMNRYQPDFVLCIGQAGGRSLTPERVAINQDDARTSDNEDNQPID
 RPIRPDGASAYFSSLPIKAMVQAIKKKDYRPLFPIRQGLLSAAI*

Description:

PYRROLIDONE-CARBOXYLATE PEPTIDASE (EC 3.4.19.3) (5-OXOPROLYL-
 PEPTIDASE). - STR EPTOCOCCUS PYOGENES.

Assembly ID: 3865378

Assembly Length: 1060bp

> 3865378 Strep Assembly -- Assembly id#3865378

CTACTTGAAACAGAACTGAAATTATACCCACTACCTCCCTGATTATCTTCAATGCTTACG
 TCTAAATAAACTTCCCCACTATTATTTAGCTTAGCAACAACCTGTTATAGTAAATAACAT
 AAAATTCACATAAATAGATTAGGGAAATCAAAGCAACTTCTAGGAATGTTTTAGCAGTCA
 CAGTGTACTTTCCAGCATCAAGCCACTATAACTCTGCACATAAAAATGGAGAAGATGGC
 CATCCTCTTCTCCAAATATTAACCTTCTTTACAAACCAACTATAGTTGACAAAGAACCTAA
 AATCAATTGATAACACGAGGTCAGGTCCGTCAACTCTTTCAACTGAAGCCCTGTCAACTC
 TTCCCATTTATCAATCTTGTATTGGAGAGAATTGCGGTGCAGATAGAGTTGCTGGGCTGT
 TTAAGTGAGAACAGCACTATTTTCCCAAAGAGAGAGAATGATTTCTGAATCTGATCTTG
 ATCCAAAATCATCTGGTGTAGACATTCTTGATTGGCTTCAAGTCCACGAGTCTTTCTCC
 CAGACTCCAAAGATAGAGCTGAGAAAAAGTATGAACACCTTGGTGACCCCTGACGCCACCA
 TGTCTTGAAACAAATCCCGCTCAGCTTTGATTAAGTCTGATAGGGCTTGATGTCCCGTCTG
 AGACCAAACCTGACCCAACATGATAGAAAGACGAAGTCCAAAGTCATACTCAACCGCTTC
 AATCGTATCACTTAAAATATCTCTTACAGAAGTGTATTTGTCTTGTTGAAGCACGAAAAC
 ATAATCCTGAGATCCGACCTGTAGCACTGTCTGACAATTCGGAAGAGTCCGCATCAT
 ATCTAGCCAAGAAGCCAGATTTTCTTGCTGAAAATAAGAAAGATGGCAATAAACCAACTG
 AATCTTTTTTAAAAAATTGCGGTGCCTGTCCCTTGCTTCAACCAGATAGGAATACCAAGG
 GTTTAGCGAACGAACCTGCTCCTGCTGGGTCAAAGGGCAACCAACTGCTTTTACGCTC
 GCTGAGCCCAGCTTCCTCCAGCAAAATCCACTGCTGAGAG

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 421 | 807 | R | 129 aa |

> 3865378-1 ORF translation from 421-807, direction R
 VLQVGSQDYVFLQQDKYTSVRDILSDTIEAVEYDFGLRLSIMLGQVWSQTGHQALSDLI
 KAERDLFKTWWRQGHQGVHTFSQLYLWLSLGERLVDLKPKECLHQMILDQDQIQEIILSL
 WENSAVLT*

Description:

unknown

Assembly ID: 3865470
 Assembly Length: 895bp

> 3865470 Strep Assembly -- Assembly id#3865470

ATTTTAGACTTTGATGACAATCCTCAGGCGGTTATCATGCCCAATCACGAGGGGCTGGAA
 TTGCAGTTGCCAAAGAAGTGTGTTTATGCATTTTATAGGTGAGGAGATCTGACCGCTATGC
 AAGGGAAGTAGGGGCGGATTGTGTGGCGAATTCGTTTCTGCTACCAAGACCTATCCAGT
 CTCTTTCATCAACTACAAGGGTGAGGAGGTCTGTCTGGATCAGGCTCCTGCTGGCTCCGC
 TCCAGCAGCCCAGTTTATGGATGGGTGATTGGCTATGGTGTGGAGCAGCTTATCTCTAC
 TGGGACCTGTGGTGTCTAGCTGATATAGAGGAAAATGCCTTCTAGTCCCTGTTTCGCGC
 TTTGCGAGATGAGGGAGCCAGTTACCACCTATGTGGCACCTTGTCGTTATATGGAAATGCA
 GCCAGAGGCTATTGCTGCTATTGAGGAAGTTTGGGAAGACAGAGGGATTCCCTTATGAAGA
 AGTCATGACCTGGACGACAGACGGTTTTTACCGAGAAACGGCTGAAAAGGTGGCTTATCG
 TAAGGAAGAAGGCTGTGCTGTTGTGGAGATGGAGTGTCTGCTCTTGCGGCAGTAGCTCA
 ATTGCGTGGGGTTCTCTGGGGTGAATTGTTGTTTACAGCAAATCTCTAGCGGACTTGGA
 CCAGTACAACAGTCGTGACTGGGGCTCGGAACCTTTTAATAAGGCGCTAAAACCTGAGTTT
 AGCAAGTGTCACACCTTTAGTTGTACTGGCAAAGGATTTGTTTTATCATAAAATGTCT
 AGCTCATACTTTTCAAAAATATGTTTAAACGAAAGTCACCTTCCTCTTGTCTTAAGCATGT
 TTGAAGTTGGGAAAAATCTTTAAATCAGAAAAACGTATCATATCAGGTTGATGA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 98 | 742 | F | 215 aa |

> 3865470-1 ORF translation from 98-742, direction F

VRRSDRYAREVGADCVGEFVSATKTYPVVSFINYKGEEVCLDQAPAGSAPAAQFMDGLIGY
 GVEQLISTGTCGVLADIEENAFVLPVRALRDEGASYHYVAPCRYMEMQPEAIAAIEEVLE
 DRGIPYEEVMTWTTDGFYRETAEKVAYRKEEGCAVVEMECSALAAVAQLRGVLWGELLFT
 ANSLADLDQYNSRDWGSEPFNKALKLSLASVHHL*

Description:

unknown

Assembly ID: 3865632
 Assembly Length: 645bp

> 3865632 Strep Assembly -- Assembly id#3865632

AGGGCTGTCAAGCTTGGTTAGAACGTTTAGAAAAGGAGAGTTAAGGTGGAAAATCTTACG
 AATTTTTACGAAAAGTATCGTGTCTATCTGACTCGTCCACGTTTAGAGCTTTTGGCAGTA
 GTTACCATTGTTTTANGNGCTGTACTCGTCTTTTTTCTAAATATCCAGGAAAAGGTGTC
 TTAAAACTCGATAATGGAACGATTGTTTATGATGGCAGTCTTGTCCTGGTAAAATGAAT

GGCCAAGGTACCATTTACCTTCCAAAATGGAGACCAATATACAGGTGGCTTCAACAATGGA
 GCCTTCAACGGAAAAGGTACCTTTCAATCTAAAGAAGGCTGGACCTACGAAGGTGATTTT
 GTAAATGGTCAGGCTGAAGGAAAAGGGAACTAACAACAGAACAGAAGTCGTTTATGAA
 GGAACTTTTTAAACAAGGCGTTTTTCAACAAAAATAAAGCCTCCTTATCAAAGGAGGTATT
 ATTAGAATTACAAGGTAAGCGTTTACCTGTAAATCCCTTTCTTTCCAAATCCCTCTTCCA
 AGCAAGTTTGTGAAATAAAAAATATTTGAAATAAATTTACAAACTTCAAAGATAAAACC
 TGATAAGAAAAGAAAATGAGAAAAGTTTCGCAAGAGTTTAAAAAT

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 46 | 456 | F | 137 aa |

> 3865632-1 ORF translation from 46-456, direction F
 VENLTNIFYEKYRVYLTRPRLELLAVVTIVLXAVLVFFLNIPGKGVCLKLDNGTIVYDGS
 LGSLV
 RGKMNQGQGTITFQNGDQYTGGFNNGAFNGKGTFFQSKEGWTYEGDFVNGQAEGKGLTTEQ
 EVVYEGTFTKQGVFQQK*

Description:

unknown

Assembly ID: 3865710

Assembly Length: 572bp

> 3865710 Strep Assembly -- Assembly id#3865710

GAGATCTGTCTTGACACCAAAGTGTGGAGTACGCCAGCTAATTCAACGGCGATATAACC
 AGCGCCTAGCAATCGCAATTGACTCTGGAAGTTCTTCCCAGGCAAATACATCAGAGAAG
 GCCACCTAGCTCAGCACCAGGAATATTAGGAATACTTGGATGGGCACCTGTAGCAATCAC
 GATATGTCTAGCACGAATCAGTTCACCATTTACGCTTACAGTATGAGAATCTACAAATTC
 AGCATGACCTTCAATCAAGTCTACACCGTTGCGTTTAAACTACCATCATAGAGAAGAAC
 GAGCGCGATCAATGTAGGCTTCACGATTGCGACGTAGGGTTGCAAAGTTAAAGTTAAGAT
 CAGTAGTCTCAAAGCCGTAGTCTCCTCCAAATTGATGGAAAGTCTCAGCGATTGCGCCC
 CGCTACCACATGATTCTTTTAGGAACACAACCGACGTTGACACAGGTTCCACCTAATTC
 TTTTCCTCAATAACGGCTGCTTTGGCTCCATGTTCCCAGCACGGTTCATGGTAGCGATCC
 TCCGCTACCTCCACGATAGCAATGATATCATA

ORF Predictions:

| ORF # | Start | End | Direction | Length |
|-------|-------|-----|-----------|--------|
| 1 | 287 | 448 | R | 54 aa |

> 3865710-1 ORF translation from 287-448, direction R
 VFLKESCGSGAQIAETFHQFGGDYGFETDLNFNFATLRRNREAYIDRARSSL*

Description:

glutathione reductase (NADPH) (EC 1.6.4.2) - *Streptococcus thermophilus*

Provided in Table 2 is information on the direction of the ORF (forward or reverse) for each polynucleotide in Table 1. Also listed for each ORF is its start and stop codon positions (refer to the columns containing nucleotide code labeled "Start" and "Stop"). The triplet codon sequence for each start and stop codon is also shown. These codons may be shown in the sense orientation or antisense orientation, such as GTG and CAC, respectively, for start codons. The "Length" column discloses the length of each polynucleotide assembly. The direction of translation on the polynucleotide depicted is denoted by and "Forward" for forward or and "Reverse" for reverse (or being on the opposite strand from the one depicted). As indicated above, the "Assembly ID" number is a unique identifier assigned to each ORF of Table 1 and allows a correlation between the data in Tables 1 and 2.

TABLE 2

| Assembly ID | Start | Stop | Start | Stop | Length | Direction |
|-------------|-------|------|-------|------|--------|-----------|
| 3049156 | ~CAC | TCA~ | 236 | 385 | 50 | Reverse |
| 3049862 | GTG | TGA | 383 | 526 | 48 | Forward |
| 3112810 | ~CAC | TTA~ | 601 | 804 | 68 | Reverse |
| 3112866 | ~CAC | TTA~ | 220 | 513 | 98 | Reverse |
| 3113664 | GTG | TAA | 165 | 392 | 76 | Forward |

| Assembly ID | Start | Stop | Start | Stop | Length | Direction |
|----------------|-------|------|-------|------|--------|-----------|
| 3113716 | ~CAC | TTA~ | 94 | 291 | 66 | Reverse |
| 3174176 | GTG | TAA | 139 | 543 | 135 | Forward |
| 3174186 | GTG | TAG | 83 | 283 | 67 | Forward |
| 3174374 | GTG | TGA | 154 | 294 | 47 | Forward |
| 3174972 | ~CAC | TTA~ | 169 | 678 | 170 | Reverse |
| 3175138 | ~CAC | TCA~ | 79 | 945 | 289 | Reverse |
| 3175860 | GTG | TAA | 51 | 251 | 67 | Forward |
| 3175918 | GTG | TGA | 212 | 535 | 108 | Forward |
| 3811220 | ~CAC | CTA~ | 316 | 873 | 186 | Reverse |
| 3811436 | ~CAC | TTA~ | 1164 | 1511 | 116 | Reverse |
| 3811984 | GTG | TGA | 134 | 454 | 107 | Forward |
| 3857228 | ~CAC | TCA~ | 1141 | 1356 | 72 | Reverse |
| 3857842 | GTG | TAA | 45 | 341 | 99 | Forward |
| 3857996 | GTG | TAA | 58 | 456 | 133 | Forward |
| 3858236 | ~CAC | CTA~ | 1 | 261 | 87 | Reverse |
| 3858264 | ~CAC | TCA~ | 439 | 1365 | 309 | Reverse |
| 3858610 | ~CAC | TTA~ | 374 | 949 | 192 | Reverse |
| 3858716 | ~CAC | CTA~ | 238 | 402 | 55 | Reverse |
| 3859124 | ~CAC | CTA~ | 73 | 453 | 127 | Reverse |
| 3859244 | ~CAC | TTA~ | 310 | 462 | 51 | Reverse |
| 3859250 | ~CAC | CTA~ | 244 | 402 | 53 | Reverse |
| 3859588 | ~CAC | TTA~ | 102 | 443 | 114 | Reverse |
| 3859774 | ~CAC | CTA~ | 9 | 131 | 41 | Reverse |
| 3860140 | GTG | TAA | 302 | 511 | 70 | Forward |
| 3860140 | GTG | TAA | 605 | 856 | 84 | Forward |
| 3860206 | ~CAC | TTA~ | 898 | 1056 | 53 | Reverse |
| 3860270 | GTG | TAG | 346 | 966 | 207 | Forward |
| 3860438 | GTG | TAG | 1 | 276 | 92 | Forward |
| 3860438 | GTG | TGA | 460 | 1128 | 223 | Forward |
| 3860544 | GTG | TAA | 222 | 689 | 156 | Forward |
| 3860558 | ~CAC | TTA~ | 717 | 1376 | 220 | Reverse |
| 3860568 | GTG | TAA | 1040 | 1291 | 84 | Forward |
| 3860582 | GTG | TGA | 356 | 1027 | 224 | Forward |
| 3860724 | GTG | TGA | 139 | 498 | 120 | Forward |

| Assembly ID | Start | Stop | Start | Stop | Length | Direction |
|-------------|-------|------|-------|------|--------|-----------|
| 3860724 | GTG | TGA | 686 | 1024 | 113 | Forward |
| 3860858 | GTG | TAG | 610 | 807 | 66 | Forward |
| 3860890 | GTG | TAG | 397 | 486 | 30 | Forward |
| 3860952 | ~CAC | TTA~ | 449 | 715 | 89 | Reverse |
| 3860962 | ~CAC | TTA~ | 152 | 646 | 165 | Reverse |
| 3861268 | ~CAC | TTA~ | 457 | 645 | 63 | Reverse |
| 3861270 | ~CAC | TTA~ | 627 | 824 | 66 | Reverse |
| 3861288 | ~CAC | CTA~ | 357 | 572 | 72 | Reverse |
| 3861306 | GTG | TAA | 717 | 1208 | 164 | Forward |
| 3861306 | GTG | TAA | 1201 | 1410 | 70 | Forward |
| 3861334 | GTG | TAA | 76 | 975 | 300 | Forward |
| 3864148 | GTG | TAG | 212 | 940 | 243 | Forward |
| 3864148 | GTG | TAA | 1202 | 1753 | 184 | Forward |
| 3864148 | GTG | TAA | 2750 | 3037 | 96 | Forward |
| 3864172 | GTG | TAG | 311 | 862 | 184 | Forward |
| 3864180 | ~CAC | TTA~ | 930 | 1616 | 229 | Reverse |
| 3864184 | GTG | TGA | 197 | 670 | 158 | Forward |
| 3864184 | GTG | TAA | 612 | 1304 | 231 | Forward |
| 3864194 | ~CAC | CTA~ | 1084 | 1380 | 99 | Reverse |
| 3864338 | GTG | TGA | 552 | 1100 | 183 | Forward |
| 3864360 | GTG | TAA | 47 | 1078 | 344 | Forward |
| 3864388 | GTG | TGA | 1239 | 1586 | 116 | Forward |
| 3864406 | ~CAC | TTA~ | 263 | 958 | 232 | Reverse |
| 3864452 | ~CAC | TCA~ | 1079 | 1201 | 41 | Reverse |
| 3864458 | GTG | TAA | 797 | 1105 | 103 | Forward |
| 3864458 | GTG | TGA | 1179 | 1391 | 71 | Forward |
| 3864474 | ~CAC | CTA~ | 68 | 247 | 60 | Reverse |
| 3864474 | ~CAC | TTA~ | 644 | 1528 | 295 | Reverse |
| 3864510 | ~CAC | TTA~ | 1164 | 1640 | 159 | Reverse |
| 3864526 | ~CAC | TTA~ | 845 | 1660 | 272 | Reverse |
| 3864548 | GTG | TGA | 687 | 1055 | 123 | Forward |
| 3864548 | GTG | TAA | 979 | 1932 | 318 | Forward |
| 3864582 | ~CAC | TTA~ | 317 | 550 | 78 | Reverse |
| 3864604 | ~CAC | CTA~ | 1 | 141 | 47 | Reverse |
| 3864604 | ~CAC | CTA~ | 1513 | 1803 | 97 | Reverse |
| 3864610 | GTG | TAA | 427 | 1305 | 293 | Forward |
| 3864716 | GTG | TAA | 57 | 272 | 72 | Forward |
| 3864718 | GTG | TGA | 77 | 1474 | 466 | Forward |
| 3864802 | ~CAC | TTA~ | 92 | 550 | 153 | Reverse |

| Assembly ID | Start | Stop | Start | Stop | Length | Direction |
|-------------|-------|------|-------|------|--------|-----------|
| 3864854 | ~CAC | CTA~ | 324 | 548 | 75 | Reverse |
| 3864862 | ~CAC | CTA~ | 431 | 1003 | 191 | Reverse |
| 3864888 | ~CAC | TTA~ | 10 | 657 | 216 | Reverse |
| 3864898 | GTG | TAA | 130 | 1029 | 300 | Forward |
| 3864938 | GTG | TGA | 883 | 1326 | 148 | Forward |
| 3864956 | GTG | TAA | 1030 | 1251 | 74 | Forward |
| 3864958 | ~CAC | TCA~ | 1427 | 1711 | 95 | Reverse |
| 3865022 | ~CAC | TCA~ | 279 | 1271 | 331 | Reverse |
| 3865036 | GTG | TAG | 79 | 492 | 138 | Forward |
| 3865054 | ~CAC | TCA~ | 302 | 793 | 164 | Reverse |
| 3865102 | ~CAC | CTA~ | 27 | 731 | 235 | Reverse |
| 3865156 | ~CAC | TTA~ | 416 | 808 | 131 | Reverse |
| 3865160 | GTG | TAA | 136 | 375 | 80 | Forward |
| 3865172 | ~CAC | TTA~ | 731 | 1123 | 131 | Reverse |
| 3865228 | GTG | TAA | 197 | 286 | 30 | Forward |
| 3865230 | GTG | TGA | 272 | 586 | 105 | Forward |
| 3865378 | ~CAC | TTA~ | 421 | 807 | 129 | Reverse |
| 3865470 | GTG | TAG | 98 | 742 | 215 | Forward |
| 3865632 | GTG | TAA | 46 | 456 | 137 | Forward |
| 3865710 | ~CAC | TCA~ | 287 | 448 | 54 | Reverse |

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1

Isolation of DNA coding for a virulence gene in *Streptococcus pneumoniae*

As mentioned above each of the DNAs disclosed herein by virtue of the fact that it includes an intact open reading frame is useful to a greater or lesser extent as a screen for identifying antimicrobial compounds. A useful approach for selecting the preferred DNA sequences for screen development is evaluation by insertion-duplication mutagenesis. This system disclosed by Morrison et al., *J. Bacteriol.* 159:870 (1984), is applied as follows.

Briefly, random fragments of *Streptococcus pneumoniae*, strain 0100993 DNA are generated enzymatically (by restriction endonuclease digestion) or physically (by sonication based shearing) followed by gel fractionation and end repair employing T4 DNA

polymerase. It is preferred that the DNA fragments so produced are in the range of 200-400 base pairs, a size sufficient to ensure homologous recombination and to insure a representative library in *E.coli*. The fragments are then inserted into appropriately tagged plasmids as described in Hensel et al., Science 269: 400-403(1995). Although a number of plasmids can be used for this purpose, a particularly useful plasmid is pJDC9 described by Pearce et al., Mol. Microbiol. 9:1037 (1993) which carries the *erm* gene facilitating erythromycin selection in either *E. coli* or *S. pneumoniae* previously modified by incorporation of DNA sequence tags into one of the polylinker cloning sites. The tagged plasmids are introduced into the appropriate *S. pneumoniae* strain selected, *inter alia*, on the basis of serotype and virulence in a murine model of pneumococcal pneumonia.

It is appreciated that a seventeen amino acid competence factor exists (Havastein et al., Proc. Nat'l. Acad. Sci. USA 92:11140-44 (1995)) and may be usefully employed in this protocol to increase the transformation frequencies. A proportion of transformants are analysed to verify homologous integration and as a check on stability. Unwanted levels of reversion are minimized because the duplicated regions will be short (200-400 bp), however if significant reversion rates are encountered they may be modulated by maintaining antibiotic selection during the growth of the transformants in culture and/or during growth in the animal.

The *S. pneumoniae* transformants are pooled for inoculation into mice, *eg.*, Swiss and/or C57B1/6. Preliminary experiments are conducted to establish the optimum complexity of the pools and level of inoculum. A particularly useful model has been described by Veber et al. (J. Antimicrobiol. Chemother. 32:432 (1993) in which 10^5 cfu inocula sizes are introduced by mouth to the trachea. Strain differences are observed with respect to onset of disease *e.g.*, 3-4 days for Swiss mice and 8-10 days for C57B1/6. Infection yields in the lungs approach 10^8 cfu/lung. IP administration is also possible when genes mediating blood stream infection are evaluated. Following optimization of parameters of the infection model, the mutant bank normally comprising several thousand strains is subjected to the virulence test. Mutants with attenuated virulence are identified by hybridization analysis using the labelled tags from the "input" and "recovered" pools as probes as described in Hensel et al., Science 269: 400-403(1995). *S. pneumoniae* DNA is colony blotted or dot blotted, DNA flanking the integrated plasmid is cloned by plasmid rescue in *E. coli* (Morrison et al., J. Bacteriol. 159:870 (1984)) and sequenced. Following sequencing, the DNA is compared to the nucleotide sequences given herein and the appropriate ORF is identified and function confirmed for example by knock-out studies.

Expression vectors providing the selected protein are prepared and the protein is configured — in an appropriate screen for the identification of anti-microbial agents. Alternatively, genomic DNA libraries are probed with restriction fragments flanking the integrated plasmid to isolate full-length cloned virulence genes whose function can be confirmed by "knock-out" studies or other methods, which are then expressed and incorporated into a screen as described above.

What is claimed is 1. An isolated polynucleotide comprising a polynucleotide — sequence selected from the group consisting of:

- (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence of Table 1;
- (b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;
- (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence of Table 1;
- (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
- (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

- 2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
- 3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
- 4. The polynucleotide of Claim 2 comprising the nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in Table 1.
- 5. The polynucleotide of Claim 2 which encodes a polypeptide comprising an amino acid sequence sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 6. A vector comprising the polynucleotide of Claim 1.
- 7. A host cell comprising the vector of Claim 6.
- 8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.
- 9. A process for producing a polypeptide or fragment comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide or fragment.
- 10. A polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 11. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 12. An antibody against the polypeptide of claim 10.

13. An antagonist or agonist of the activity or expression of the polypeptide of claim 10.
14. A method for the treatment or prevention of disease of an individual comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 10.
15. A method for the treatment of an individual having need to inhibit a bacterial polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 13.
16. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 10 in an individual comprising:
- (a) determining a nucleic acid sequence encoding said polypeptide, and/or
 - (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.
17. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 10 comprising:
- contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;
 - and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.
18. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of claim 10, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.
19. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of a polypeptide of claim 10, or fragment or a variant thereof, for expressing said polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.
20. A polynucleotide comprising a polynucleotide sequence selected from the group consisting of the the first ten polynucleotides sequences from the top of Table 1.

21. A polypeptide comprising a polypeptide encoded by the polynucleotide of claim 20.

22. The isolated polynucleotide of claim 1 wherein said nucleotide is selected from the group consisting of:

(a) a polynucleotide having at least a 90% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 90% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 90% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

23. The isolated polynucleotide of claim 1 selected from the group consisting of:

(a) a polynucleotide having at least a 95% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 95% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

24. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*;

(b) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and

(c) a polynucleotide which is complementary to the polynucleotide of (a) or (b).

25. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

26. An isolated nucleic acid encoding one of the amino acid sequences of Claim 1 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

27. Recombinant vectors comprising the nucleic acid sequences of Claim 26 and host cells transformed or transfected therewith.

28. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 1 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

29. Antimicrobial compounds identified by the method of Claim 28.

30. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

31. An isolated nucleic acid encoding one of the amino acid sequences of Claim 30 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

32. Recombinant vectors comprising the nucleic acid sequences of Claim 31 and host cells transformed or transfected therewith.

33. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 30 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

34. Antimicrobial compounds identified by the method of Claim 33.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CA, EMBASE, WPIDS

terms: Streptococcus, pneumoniae, dna, polypeptide, treat, diagnose

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|------------------------|
| X | US 5,476,929 A (BRILES ET AL) 19 December 1995, see entire document | 1-5, 20, 22-24, 26, 31 |
| A | SEVIER et al. Monoclonal Antibodies in Clinical Immunology. Clinical Chemistry. 1981, Vol. 27, No. 11, pages 1797-1806, see entire document | 1-34 |
| A | US 4,601,980 A (GOEDEL ET AL) 22 July 1986, see entire document. | 1-34 |
| A | US 5,474,905 A (TAI ET AL) 12 December 1995, see entire document. | 1-34 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

| | |
|---|--|
| * Special categories of cited documents: | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| *A* document defining the general state of the art which is not considered to be of particular relevance | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| *E* earlier document published on or after the international filing date | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *A* document member of the same patent family |
| *O* document referring to an oral disclosure, use, exhibition or other means | |
| *P* document published prior to the international filing date but later than the priority date claimed | |

Date of the actual completion of the international search

21 JANUARY 1998

Date of mailing of the international search report

20 FEB 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MARK NAVARRO

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

A61K 38/00, 39/00, 39/395, 39/40; C07H 21/04; C07K 1/00; C12N 15/00; C12P 21/06; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)